

利用觀測環境參數與機器學習預報大台北都會區的午後雷陣雨

**Prediction of Afternoon Thunderstorm  
in the Taipei Metropolitan Area using Observations of Environmental  
Parameters and Machine Learning**

楊天瑞 Tien-Jui Yang

指導教授：王重傑 Prof. Chung-Chieh Wang

國立臺灣師範大學地球科學系

暑期學生專題研究計畫書面報告

Department of Earth Sciences National Taiwan Normal University

Earth Sciences Summer Student Program Report

2021 年 7 月 - 8 月

## ● 摘要：

近年來，人工智慧與機器學習的快速發展，可以應用於各種科學領域來幫助解決問題，在天氣預報上也有其可應用之處。機器學習的基本概念，是利用類神經網路來自我學習，使用者給予許多個案的大量相關資料以及各個個案的結果，機器以這些資料訓練，從中得出各種資料的重要性與之間的關聯，以使訓練的結果能與正確的結果最為一致。因此，本研究旨在應用機器學習的方法，由觀測的環境參數來預報大台北都會區的午後雷陣雨。研究發現測站資料與探空資料的預報度最高為 0.7915，且基隆測站對於大台北都會區是否發生午後對流有支配性的預報能力。

## ● 前言：

台灣夏季在弱綜觀條件下，常在午後發生雷陣雨的現象。此類短延時的劇烈強降雨，若發生於北部的大台北都會區，常因排水系統宣洩不及，造成局部地區淹水、導致交通癱瘓、財物損失、以及生活上的不便。在有利的環境條件下，主要由熱力過程驅動的午後雷陣雨，時常發展迅速而令人措手不及，在預報上也十分具有挑戰性。本研究應用機器學習的方法，由觀測的環境參數來預報大台北都會區暖季(2018-2019 年 5-9 月)在弱綜觀環境下的午後對流。曾有學者(LIN et al. 2012)利用模糊邏輯(fuzzy logic)概念發展北台灣地區 TSA 的客觀潛勢預報方法(Afternoon Thunderstorm Occurrence Potential forecast using Fuzzy Logic, ATOPFLO) 有效地整合對流前環境特徵，並得到不錯準確的預報度。此外，海陸風或上升與下降氣流對於邊界層的濕化，可提供適合發生午後對流有利條件(Johnson and Bresch 1991; Chen and Li 1995)。因此，複雜的台灣地形地區，其地形效應在降雨分布的變化也具有相當程度的影響。

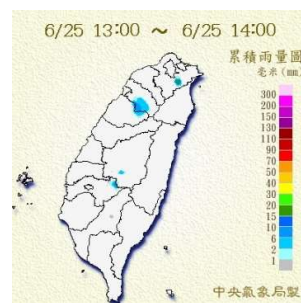
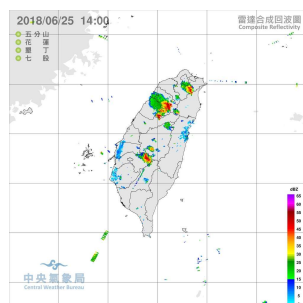
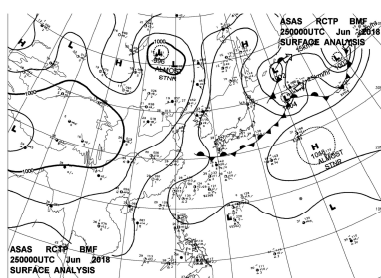
## ● 資料與分析方法：

本研究大致的內容與進程序如下：首先，我們會辨認過去數個夏季期間，在弱綜觀條件下的北部地區午後對流個案，包括雷雨發生日與未發生日。隨後，會收集每個個案在發生前一段時間的觀測環境參數，是我們認為會影響當日午後對流發生與否的參數。例如：板橋、花蓮、馬公、東沙等探空觀測(0800 LST) 所得參數 (如：地面溫度與露點、各近地層溫度與露點、對流可用位能、對流抑制能、LCL 高度、LFC 高度、CCL 高度、低層降溫率、對流指數...等)、幾個特定地面氣象站 (或自動站) 的逐時資料 (溫度、濕度、風向、風速) 等，提供輸入給機器學習訓練，而結果則是這些個案日的午後對流結果。因

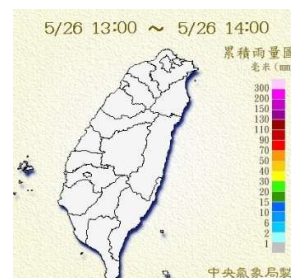
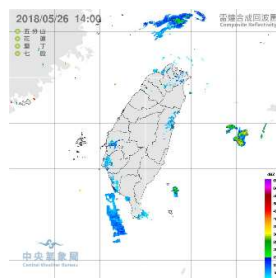
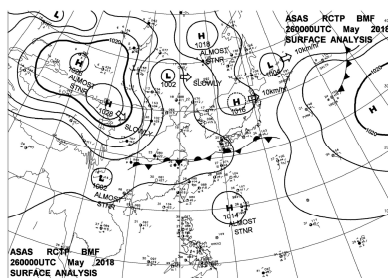
此，輸入的資料也要包括午後對流未發生的個案，如此機器學習才能分辨有無午後對流。因此機器學習的研究步驟分為三步驟：辨認有無午後對流個案、蒐集個案天數的觀測氣象參數並分析參數與是否降雨之相關係數、進行機器學習。

### (一) 資料與個案辨認：

弱綜觀環境下台灣北部發展的午後對流，比對 2018、2019 年 5 到 9 月，共 10 個月的大氣水文資料庫天氣圖，台灣周圍需為乾淨環境場(北緯 20-28 度、東經 118-124 度內無鋒面；北緯 19-28 度、東經 116-126 度內無颱風)。這些弱綜觀環境天數下，若下午 12:00-21:00 間雷達迴波大於 40dBZ 且回波範圍大於  $10\text{km}^2$ ，降水強度大於每小時 5 毫米則定義此天為有午後對流個案，10 個月中共有 31 天；反之，若下午 12:00-21:00 間很少或沒有雷達迴波且無降水則被分類為無午後對流個案，共有 71 天。



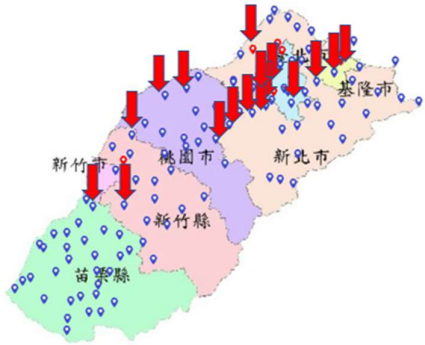
(圖一)、(圖二)、(圖三)為有午後對流個案的例子，可知當天台灣附近為弱綜觀環境場，且大台北都會帶在下午 12:00-21:00 雷達迴波大於 40dBZ 且回波範圍大於  $10\text{km}^2$ ，降水強度大於每小時 5 毫米。



(圖四)、(圖五)、(圖六)為無午後對流個案的例子，當天台灣附近為弱綜觀環境場，而大台北都會帶在下午 12:00-21:00 幾乎沒有雷達迴波，且當天無降水。

### (二) 個案資料蒐集：

利用 CWB 的測站資料以及探空資料，分別蒐集北台灣沿河谷地形共 18 個測站(如圖七)：基隆、台北、淡水、七堵、新莊、五股、汐止、文山、中和、永和、土城、三峽、大溪永福、大園、觀音、湖口、頭份、峨眉；以及板橋、花蓮、馬公、東沙四個測站探空資料。而測站資料為 8:00~12:00(LST)小時資料，包含氣壓、氣溫、相對濕度、風向、風速等 5 種資料；探空資料則是蒐集每日上午 8:00(00Z)的露點溫度、K.index 等 13 種探空資料。



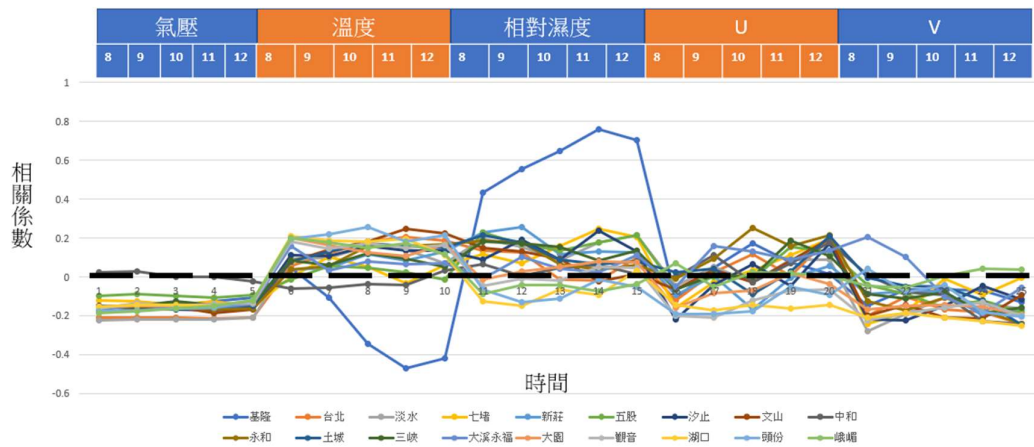
Hourly Station Data 8:00~12:00(LST)	Sounding (00 Z) :
Air pressure Temperature Relative humidity Wind direction Wind speed(u、v)	PO、TO、Td0、LCL、 CCL、LFC、K.index、 total lifted index、 showalter index、sweat index、CAPE、CIN
North Taiwan Along River Bank(18 Weather Stations)	Banqiao、Hualien、 Magong、Dongsha

(圖七) 北台灣沿河谷地形共 18 個觀測站分布。

(表一)測站資料與探空資料。

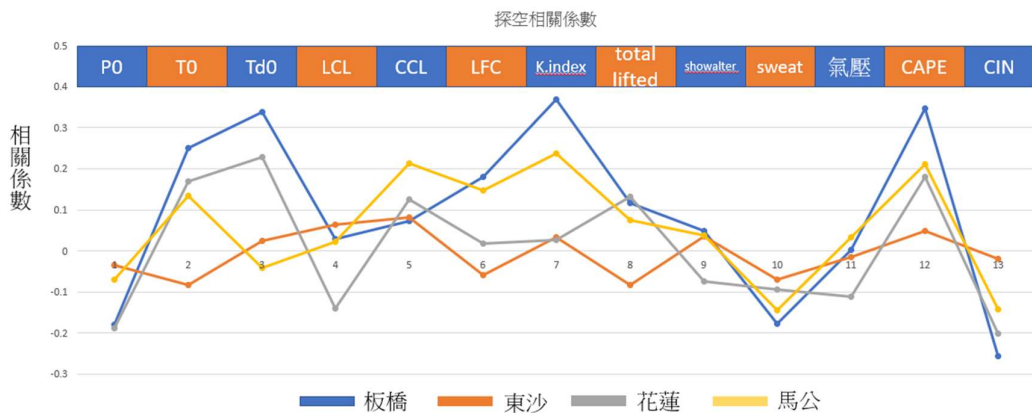
蒐集完資料後，由於蒐集的大量資料不一定都能對目標(預測是否有午後對流發生)有效預測，因此先分別分析測站資料、探空資料與是否有午後對流發生的相關係數(有午後個案天數當作 1，無午後個案計做 0)，將相關係數差的資料排除，以提高機器學習的效率。測站資料(如下圖八，18 個測站氣象參數與是否發生午後對流之相關係數，縱軸為相關係數，橫軸依序則為 8:00~12:00 氣壓、溫度、相對溼度、U 方向風、V 方向風)的相關係數大多在正負 0.2 之間，這樣的相關性在機器學習中是可以接受的相關程度，意即這些氣象參數有預報的價值。此外值得注意的是基隆測站(圖八中藍色突出線)在溫度、相對濕度的相關係數由 8~12 點之間快速上升，而最高點出現在上午 11 點。而探空資料(如圖九)的相關係數則落在-0.2~0.4 之間，比測站資料普遍高一些，其中露點溫度、K.index、CAPE 與 CIN 係數較高，此外測站中又以板橋花蓮的相關係數較高，推測因為接近對流發生區域，因此更能反應出和環境場資訊的相關性。

## Correlation Coefficient (weather station)



(圖八) 18 個測站氣象參數與是否發生午後對流之相關係數，橫軸依序則為 8:00~12:00 氣壓、溫度、相對溼度、U 方向風、V 方向風。

## Correlation Coefficient (Sounding)

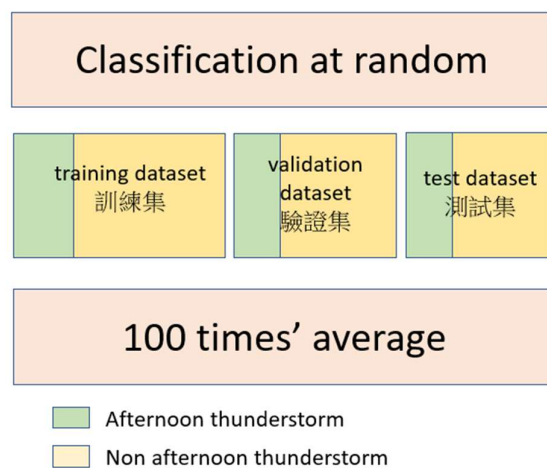


(圖九)板橋、花蓮、東沙、馬公探空資料與是否發生對流之相關係數，縱軸為相關係數，橫軸為探空參數。測站中以板橋花蓮的相關係數較高，推測因為接近對流發生區域，因此更能反應出環境場的關聯。

### (三) 機器學習

機器學習的基本概念是利用類神經網路來自我學習，給予大量相關資料以及結果(是否有午後對流發生)，以這些大量資料訓練，並以訓練的結果能與正確的結果比較找出最佳的訓練途徑。將所有資料「按比例」且「隨機」分為三類(如圖十)：訓練集(training dataset)、驗證集(validation

dataset)、測試集(test dataset)，而每種 dataset 中有發生午後對流個案天數與無發生午後對流個案天數皆維持 3：7 的比例(有午後對流發生：無午後對流發生)。將訓練集(training dataset)放入機器學習模型中(如圖十一，機器學習模型共有 5 層，並由 Sigmoid function 以及 ReLU function 交互組成，而每層皆有 512 個節點)，進行學習。同時驗證集(validation dataset)的腳色是計算不同疊帶次數的 loss 值(如圖十二)，並找到最小的 loss 值(loss 越小代表預測能力越好)，則此時的疊帶次數對於訓練集(training dataset)在機器模型中的訓練結果能有最佳的預報度。訓練好模型後再將測試集(test dataset)資料套用到模型中進行結果比對，得出一個預報度。重複上述步驟 100 次並且將這 100 次的預報度取平均值，得出有代表性的預報度。

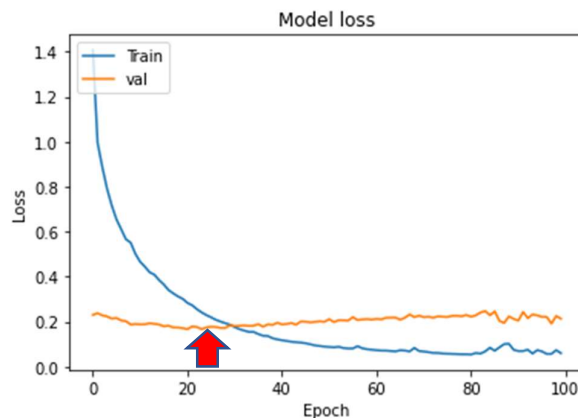


(圖十) 訓練集(training dataset)、驗證集(validation dataset)、測試集(test dataset)，而每種 dataset 中有發生午後對流個案天數與無發生午後對流個案天數皆維持 3：7 的比例(有午後對流發生：無午後對流發生)，將訓練集(training dataset)放入圖十一的機器學習模型中。再依據驗證集(validation dataset)計算不同疊帶次數的 loss 值，找到最小 loss 值，依據 loss 值的蝶帶次數，訓練好模型後再將測試集(test dataset)資料套用到模型中進行結果比對，得出一個預報度。重複上述步驟 100 次並且將這 100 次的預報度取平均值，得出有代表性的預報度。



Fully Connected Layer	512 nodes	Sigmoid
Fully Connected Layer	512 nodes	ReLU
Fully Connected Layer	512 nodes	Sigmoid
Fully Connected Layer	512 nodes	ReLU
Output layer	512 nodes	Sigmoid

(圖十一)機器學習模型共有 5 層，並由 Sigmoid function 以及 ReLU function 交互組成，而每層皆有 512 個節點，進行學習。



(圖十二)同時驗證集(validation dataset)的腳色是計算不同疊帶次數的 loss 值，並找到最小的 loss 值(loss 越小代表預測能力越好)，則此時的疊帶次數對於訓練集(training dataset)在機器模型中的訓練結果能有最佳的預報度。

## ● 結果與討論

(一) 資料分成三種實驗方式：

將資料分成三種實驗方式：(1)測站資料與探空資料(2)只有測站資料(3)只有探空資料。分別進行機器學習的預報結果如下表二，(1)測站資料與探空資料的預報度最高為 0.7915，其次是(2)只有測站資料為 0.769，最低則是(3)只有探空資料為 0.6965。這樣的結果是符合預期的，因為一般來說投入越多的資料到模型中，越能有更高的預報準確度，而探空的資料量最

少，因此預報度為三者最低是合理的。而再將三種結果與前面提及的 ATOPFLO 預報方式 (LIN et al. 2012)相比較(如表三與圖十三)，因為測站資料是蒐集從台灣時間 8:00~12:00，因此(1)測站資料與探空資料(2)只有測站資料需要與 ATOPFLO 中的 12:00 預報度(0.78)比較；此外因為探空資料一般是台灣時間上午 8:00 施放，而大約會在上午 9:00 接收資料，因此(3)只有探空資料要與 ATOPFLO 中的上午 9:00 的預報度(0.72)相比較。比較結果發現，(1)測站資料與探空資料的預報度較 ATOPFLO 的 12:00 預報度好，而(3)只有探空資料則較 ATOPFLO 的 9:00 預報度差。

Station & Sounding	Station only	Sounding only
0.7915	0.769	0.6965

(表二) 資料分成三種實驗方式：(1)測站資料與探空資料(2)只有測站資料(3)只有探空資料。

ATOPFLO 9:00	ATOPFLO 12:00
0.72	0.78

(表三) ATOPFLO 預報方式上午九點與十二點的預報度。

時間 (LST)	Hit	Miss	False alarm	Correct rejection	POD	FAR	CSI
0800	40	5	27	46	0.889	0.403	0.556
0900	39	6	27	46	0.867	0.409	0.542
ATOPFLO 1000	39	6	25	48	0.867	0.391	0.557
1100	39	6	21	52	0.867	0.350	0.591
1200	38	7	19	54	0.844	0.333	0.594

(圖十三) ATOPFLO 預報方式 (LIN et al. 2012)。

(二)基隆測站排除之比較：

前面有提及基隆測站與是否發生午後對流相對於其他 17 個測站有較高相關係數，因此再做另一個實驗：(4)測站資料與所有探空資料，但排除基隆測站、(5)測站資料，但排除基隆測站、(6)只有基隆測站資料。結果如下圖十四，(4)測站資料與所有探空資料，但排除基隆測站的預報度為



0.7015，比(1)含基隆測站的預報度明顯下降了 0.09，而同樣的情況也發生在(5)測站資料，但排除基隆測站(0.667)與(2)含基隆測站，此外(6)只有基隆測站的預報到甚至來到 0.834，這說明了基隆測站對於大台北都會區是否發生午後對流有支配性的預報能力(海陸風或上升對流對邊界層的濕化，有利午後對流發展，因此推測基隆上午海風帶入水氣可能顯著影響大台北午後對流的發展)。

	Station & Sounding	Station only	Keelung Station only
Keelung included	0.7915	0.769	0.834
Keelung excluded	0.7015	0.667	X

(圖十四)基隆測站排除之比較：(4)測站資料與所有探空資料，但排除基隆測站、(5)測站資料，但排除基隆測站、(6)只有基隆測站資料。

● 參考文獻：

1. 林品芳，張保亮及周仲島，2012：弱綜觀環境下台灣午後對流特徵及其客觀預報。大氣科學，40，77-108。
2. Chollet, F., 2015: keras, GitHub, <https://github.com/fchollet/keras>.
3. Cybenko, G., 1989: Approximation by superpositions of a sigmoidal function. Math. Control Signals Syst., 2(4), 303–314.
4. Nair, V., and G. E. Hinton, 2010: Rectified linear units improve restricted boltzmann machines. Proceedings, the 27th International Conference on Machine Learning (ICML-10), 21-24 Jun 2010, Haifa, Israel, 8 pp.