

極短期強對流機率預報技術發展

馮智勇¹ 陳奕翰¹ 林涵芳¹ 林允才² 王溫和² 吳佳蓉² 劉郁青²
多采科技有限公司¹ 中央氣象局衛星中心²

摘要

氣象局極短期強對流機率預報系統(TANC)首要目標為針對弱綜觀環境下，0.01度(約1公里)空間格點未來60分鐘之午後對流(定義為雷達回波>35 dBz)發生可能性，現階段使用包含對流胞發生啟始位置、雷達回波外延、對流可用位能、對流抑制能、層化平均濕度、雷達回波發生頻率、雷達回波變化趨勢以及地面風輻散量等八個因子，以及美國氣象單位基於模糊邏輯理論設定的權重與隸屬函數，計算格點發生強對流的可能性(likelihood)。

考量數值模式已開始發展逐時資料同化的模擬策略且空間解析度不斷提升，而衛星觀測產品與高解析地面真實場亦持續精進，極短期強對流機率預報需可彈性納入更具解釋能力的因子並依據台灣地理特性分區提供機率預報。本研究除步利用空間分群建置複雜邏輯斯迴歸方程式並進行交叉檢驗的結果認為，採用複雜邏輯斯迴歸模型除可獲得與現行TANC系統相當的預報表現外，具有因應臺灣天氣系統生成特性而增減或更換預報因子的彈性。

關鍵字: 極短期、弱綜觀、機率預報、午後對流、邏輯斯迴歸

一、前言

暖季弱綜觀環境下，快速發展增強的午後對流時常會伴隨著閃電、雷擊與大雨，導致人們生活的不便利。相較於梅雨鋒面與颱風...等天氣系統，暖季弱綜觀午後對流預報更是極具挑戰性，原因在於不完美的模式物理與對流參數化以及目前觀測資料密度不足以解析中尺度邊界層特性。

目前氣象局極短期強對流機率預報系統(Taiwan Auto-Nowcaster, TANC)首要目標為針對弱綜觀環境下，0.01度(約1公里)空間格點未來60分鐘之午後對流(定義為雷達回波>35 dBz)發生可能性，現階段使用包含對流胞發生啟始位置、雷達回波外延、對流可用位能、對流抑制能、層化平均濕度、雷達回波發生頻率、雷達回波變化趨勢以及地面風輻散量等八個因子，透過美國氣象單位基於模糊邏輯演算法(Fuzzy logic algorithm)設定的權重與隸屬函數，計算格點發生強對流的可能性(likelihood)。

考量數值模式已開始發展逐時資料同化的模擬策略且空間解析度不斷提升，而衛星觀測產品與高解析地面真實場亦持續精進，極短期強對流機率預報需可彈性納入更具解釋能力的因子並依據台灣地理特性分區提供機率預報。首先於使用資料與研究方法中簡述本研究使用之預報因子資料以及觀測資料，並且簡介複雜邏輯斯迴歸方法、Kmeans集群分析以及機率預報統計校驗方法，第三段敘述複雜邏輯斯

迴歸對流機率預報系統發展之概況與成效，第四段以個案的方式校驗氣象局TANC可能性預報系統與複雜邏輯斯迴歸建置之對流機率預報系統，最後則為研究總結與討論。

二、使用資料與研究方法

(一)使用資料:

本研究收集由TANC系統產製之8個因子資料，分別為對流胞發生啟始位置、雷達回波外延、對流可用位能、對流抑制能、層化平均濕度、雷達回波發生頻率、雷達回波變化趨勢以及地面風輻散量，其中，對流胞發生啟始位置與雷達回波外延是以雷達觀測為基礎後製之產品，對流可用位能、對流抑制能、層化平均濕度則為氣象局RWRP模式之輸出，雷達回波發生頻率、雷達回波變化趨勢則為Lin等(2011)統計2005-2008年間雷達回波 ≥ 40 dBz的發生頻率與變化趨勢，地面風輻散量則是以地面觀測資料演算得出。使用2018年9個歷史個案皆為暖季弱綜觀環境下之午後對流個案，挑選條件為:1.5度以內無綜觀系統、1000hPa~850hPa，風力小於15Nt、天氣概況中無華南雲雨帶、南方雲系、低壓帶等字眼、距海岸線1.5度內無大片移入的回波以及個案回波大於35dBz持續時間與面積，個案日詳見表1。

(二) 複邏輯斯迴歸(Multiple Logistic Regression):

原理是利用訓練資料進行模型建置，透過如前進選擇法(forward selection)的篩選程序得到具解釋能力的預報因子並建立複迴歸模型(multiple regression)，再將各預報因子的最新輸出代入迴歸模型得到新的機率預報值。複邏輯斯迴歸方程式如下式， p 為二分(dichotomous) Y 資料發生機率值 $p(\beta)$ ， β_i 為待求係數， N_x 則為 X 變數個數。本研究採用最大概似法(max. likelihood method)推求式中各項係數。

$$p = \exp\left(\beta_0 + \sum_{i=1}^k \beta_i x_i\right) \cdot \left(1 + \exp\left(\beta_0 + \sum_{i=1}^k \beta_i x_i\right)\right)^{-1}$$

假設共有 N_{rec} 筆 XY 配對資料($N_{rec} > N_x$)供迴歸建模，則發生機率為 $p(\beta)$ 的情況下，概似函數值 $f(\beta)$ 可表示為：

$$f(\beta) = \prod_{i=1}^{N_{rec}} p^{y_i} \cdot (1-p)^{1-y_i}$$

其中，當事件發生時， $y_i = 1$ ；若事件未發生， $y_i = 0$ 。為了方便計算，採用對數概似函數值 $\ln(f(\beta))$ 為：

$$\begin{aligned} \ln(f(\beta)) &= \sum_{i=1}^{N_{rec}} \left[y_i \cdot (\beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \dots + \beta_{N_x} x_{i,N_x}) - \ln\left(1 + e^{\beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \dots + \beta_{N_x} x_{i,N_x}}\right) \right] \\ &= \sum_{i=1}^{N_{rec}} \left[y_i \cdot (\beta x_i) - \ln\left(1 + e^{\beta x_i}\right) \right] \end{aligned}$$

$$\Rightarrow F(\beta) = \frac{\partial \ln(f(\beta))}{\partial \beta_j} = \sum_{i=1}^{N_{rec}} x_{ij} \left[y_i - \frac{e^{\beta x_i}}{1 + e^{\beta x_i}} \right]$$

令上式等於零，得一非線性聯立方程組，則可利用迭代求解係數矩陣 $\hat{\beta}$ 。

(三) 前進選擇法

前進選擇法逐項增加迴歸方程式中的 X 預報因子至 N_x 個，挑選出有解釋能力之因子，其執行步驟說明如下。

1. 建立 Y 資料與各單一 X 變數的邏輯斯迴歸方程式。
2. 計算步驟 1 中各迴歸方程式係數 β_1 的卡方 (Chi-Square) 統計值，若其對應的發生機率超過使用者設定的門檻值 (p-value) 則剔除該變數進入前進選擇流程；若發生機率小於門檻值，則保留為候選變數。

3. 於步驟 2 中的候選變數中以艾凱克訊息準則 (Akaike's Information Criterion, AIC) 為指標，挑選解釋變異量最大的 X 變數為式 (1) 中的 X1 變數，得到迴歸方程式。艾凱克訊息準則的計算方式為： $AIC = 2 * n - 2 * \ln(f(\beta))$ ， n 為迴歸式使用的變數個數。
4. 於前個步驟得到的方程式增加一個 X 變數，同樣以艾凱克訊息準則為指標，保留模式解釋能力最大者以及差異小於使用者設定的寬容值 (tolerance) 為候選名單。
5. 利用已納入模式的 X 變數線性估計步驟 4 候選名單中的各個變數，而相關係數最低的 X 變數代表與模式中既有的變數共線程度最低，納入迴歸模式。
6. 重複步驟 4 與步驟 5 直到挑選出 N_x 個變數為止。

(四) Kmeans 集群分析:

Kmeans 又稱 K-means Clustering 為非監督式學習，舉例來說，給一組身高體重資料，但不說明這組資料哪些是男生或是女生，利用非監督式學習可以嘗試著區分出男生女生。本研究將 9 個午後對流個案共 196 個時間，8 個因子當作特徵，32422 個格點當作樣本，試著利用以上特徵分出午後對流的空間分布特徵，運作步驟如下：

1. 設定群數。
2. 評估各個樣本到各群心的距離，如果樣本距離第 n 個群心更近，則該樣本屬於第 n 群
3. 計算每個群中樣本的平均位置，將群心移動至該位置
4. 重複以上步驟直到各群心位置不再發生變化。

(五) 機率預報校驗方法:

使用 Performance Diagram 校驗 TANC 可能性預報與複邏輯斯迴歸對流機率預報。Performance diagram 如**錯誤! 找不到參照來源。**，橫軸為 SR (Success Ratio)，縱軸為 POD (Probability of Detection)，其式如下： $POD = \frac{HIT}{(HIT + MISS)}$

$$SR = 1 - \frac{FalseAlarm}{(HIT + FalseAlarm)}$$

POD 越高代表預報能掌握住大部分觀測有發生的事件，而 Success Ratio 越高代表預報的假警報越少。而圖中的綠線為 TS (Threat score)，橘線代表 Bias。若當預報結果出現在圖的左上角時，代表雖然能掌握住大多數觀測有發生的事件，但也有過多的假警報，整體來說有過度預報的情形。反之，若預報結果出現在右下角時，代表雖然出現假警報的次數較少，但卻漏掉許多觀測有發生的事件，因此有低估的情形。當預報

結果出現在右上角時，代表此預報結果能掌握多數觀測有發生的事件且鮮少有假警報的情況，因此當預報結果越靠近右上角代表預報表現越佳。

三、複邏輯斯迴歸模型建置與評估

本研究使用交叉驗證的方式，將9個午後對流個案逐一抽出資料集，利用剩下8個個案進行模型建置後預報抽出的個案，用以瞭解模型的可靠度以及建模方法是否適用於所有個案，表2顯示，透過複邏輯斯迴歸建置模型所選擇的預報因子順序與現行TANC系統之預報因子權重大小有很高的相關性。

透過圖1可以看到20180606、20180607、20180629、20180707這四個個案在交叉驗證中顯示模型呈現過度預報的，從ROC圖中則可以看到大多數的模型對於大於等於35dBz的事件是有區辨能力的。而20180719、20180725、20180727、20180806、20180811，這四個個案則是有較接近於對角線的可靠度。因此，我們好奇為何不同個案交叉驗證的可靠度會有如此差異，從圖2顯示上述的5個個案對流多分布在西南部地區，在交叉驗證的過程中，即便剔除欲驗證的個案，還是有足夠的對流發生資訊進入模型中，使預報表現能比其他四個對流主要發生在北部、中部或是全臺的個案還要好，因此，本研究使用Kmeans集群分析，將9個個案196個時間的8個預報因子當作特徵，空間格點當作樣本，利用每個格點上的1568個特徵對空間進行分類，圖3為空間分類後的結果，與一般午後對流降雨特性有很高的一致性。

由於空間分群建模後資料量驟減，使得該群僅有少數午後對流發生的個案導致無法建出良好的預報模型，因此，本研究將個案增加為2018年的56個暖季弱綜觀日進行分群建模與預報測試，除了桃竹苗地區(第2群)與沿海地區(第3群)的交叉驗證為過度預報的情況，有可能為2018年弱綜觀環境下於桃竹苗以及沿海地區發展的午後對流的個案數較少，因此，建模資料中有對流發生之資料量較少，導致模型預報效果不佳，其他區域結果則顯示模型有良好的可信度以及區辨能力。

四、與現行TANC可能性預報比較

本研究收集2020年6到7月間的15個弱綜觀午後對流個案進行TANC可能性預報與複邏輯斯迴歸機率預報之比較，個案列表如表2。

根據過去TANC可能性預報校驗顯示0.4為可能性預報之最佳門檻，表示可能性大於等於0.4，則可視為發生大於等於35dBz對流，而複邏輯斯迴歸對流機率預報則是在建模擬合後利用ROC曲線尋找各機率門檻下命中-假警報的最大值作為此機率預報模

型的最佳機率門檻，本研究將臺灣分成5個區域，5個分區最佳機率門檻分別為0.25、0.23、0.29、0.27、0.21，因此我們將2020年6~7月間的15個午後對流個案可能性預報、機率預報與觀測資料計算Performance Diagram。

Performance diagram (圖5)的分析顯示：TANC可能性預報過度預報(有很高的POD很低的SR)，複邏輯斯機率預報有些微的預報不足；兩者個的TS皆在0.1~0.3之間；TANC大部分的BIAS是介於2~5之間而複邏輯斯機率預報大部分是介於0.3~0.8之間。整體而言，複邏輯斯機率預報相對於TANC有較佳的預報表現。

五、總結與未來展望

本研究使用邏輯斯回歸分析衛星中心TANC系統產出的8個預報因子解釋回波強度大於35dBz的個案可知邏輯斯回歸所選進的因子解釋能力順序與TANC系統給予各預報因子權重大小排序有一致性，意味著由邏輯斯回歸所建立之對流機率預報系統應與現行之TANC系統表現相當。從交叉驗證綜合校驗中可知道，2018年弱綜觀個案發生午後對流分布區域不盡相同，因此，每個個案交叉驗證的表現不一，透過個案的雷達迴波分布圖可以知道建模需要分區域，並且從分區後的校驗中也顯示模型的預報能力有顯助的提升，最後透過2020年6到7月間的15個弱綜觀午後對流個案進行TANC可能性預報與複邏輯斯迴歸機率預報之比較，複邏輯斯機率預報相對於TANC有較佳的預報表現。

受限於現行TANC系統預報因子中之雷達回波發生頻率、雷達回波變化趨勢為固定的氣候統計值，且時間只有早上12點到晚上8、9點，因此，對流機率預報系統最早的一報只能是每日13點開始，若是對流有較早發生的時候，便無法達到提前預警的效果。後續建議從模式原始資料、衛星觀測資料、雷達回波資料、雷達回波三維風場反演資料與地面真實氣象場資料中，評估具解釋能力的預報因子而建置機率預報模型，發展一套獨立於TANC系統之外的對流機率預報作業系統。

五、參考文獻

1. Barron, J., D. Fleet, and S. Beauchemin, 1994: Performance of optical flow techniques. *Int. J. Comput. Vision*, **12**, 43–77.
2. Chang, H. L., H. Yuan, and P. L. Lin, 2012: Short-range (0-12h) PQPFs from time-lagged multimodel ensembles using LAPS. *Mon. Wea. Rev.*, **140**, 1496–1516.
3. Chang, H. L., 1997: The study on the afternoon convection in Taiwan area. M.S. thesis.

Department of Atmospheric Sciences, National Central University, Taiwan. 200 pp.

4. S.-C. Yang, H. Yuan, P. L. Lin and Y. C. Liou, 2015: Analysis of relative operating characteristic and economic value using the LAPS ensemble prediction system in Taiwan area. *Mon. Wea. Rev.*, **143**, 1833–1848.
5. Chen, Y. L., and J. Li, 1995: Characteristics of surface airflow and pressure patterns over the island of Taiwan during TAMEX. *Mon. Wea. Rev.*, **123**, 695–716.
6. Jou, B. J.-D., 1994: Mountain-originated mesoscale precipitation system in northern Taiwan: A case study 21 June 1991. *Terr. Atmos. Ocean. Sci.*, **5**, 169–197.
7. Lin, P. F., P. L. Chang, Ben J. D. Jou, J. W. Wilson and R. D. Roberts, 2011: Warm season afternoon thunderstorm characteristics under weak synoptic-scale forcing over Taiwan Island. *Wea. Forecasting*, **26**, 44–60.
8. P. L. Chang, Ben J. D. Jou, J. W. Wilson and R. D. Roberts, 2012: Objective prediction of warm season afternoon thunderstorms in northern Taiwan using a fuzzy logic approach, *Wea. Forecasting*, **27**, 1178–1197.
9. Roebber, P. J., 2009: Visualizing multiple measures of forecast quality. *Wea. Forecasting*, **24**, 601–608.

六、附圖、表

表 1 2018 年午後對流個案列表

個案日期	時間(UTC · 每 10 分鐘)
2018/06/06	06:10 ~ 10:20
2018/06/07	06:00 ~ 07:20
2018/06/29	06:00 ~ 07:40
2018/07/07	06:00 ~ 10:50
2018/07/19	06:00 ~ 10:50
2018/07/25	06:00 ~ 07:40
2018/07/27	06:00 ~ 08:50
2018/08/06	06:10 ~ 08:20
2018/08/11	06:00 ~ 08:00

表 2 2020 年午後對流個案列表

個案日期
2020/06/05
2020/06/24
2020/06/25
2020/06/27
2020/07/02
2020/07/05
2020/07/08
2020/07/11
2020/07/20
2020/07/21
2020/07/22
2020/07/23
2020/07/27
2020/07/28
2020/07/29

表 2 2018 年 9 個午後對流個案交叉驗證之因子選擇順序；
第一列為現行 TANC 系統預報因子權重

	1	2	3	4	5	6	7	8		
	RadarCu 0.315	stormInitLoc 0.135	Climate_Freq 0.21	Climate_Trend 0.09	Rh 0.15	Cape 0.02	Cin 0.03	Divergence 0.05		
個案	順序	1	2	3	4	5	6	7	8	發生時間段
20180606	3	1	4	2	5	6	7			14:10~ 18:20
20180607	1	2	3	4	5	6	7	8		13:50~ 15:20
20180629	1	3	4	5	2	6	7			13:00~ 15:40
20180707	3	1	4	5	6	8	2	7		12:30~ 18:50
20180719	1	2	3	4	7	5	6	8		13:40~ 18:50
20180725	1	3	4	5	2	6	7	8		13:00~ 15:40
20180727	1	3	4	5	2	6	7	8		12:20~ 16:50
20180806	1	5	2	3	4	7	6	8		13:50~ 16:20
20180811	1	3	4	2	5	6	7			13:30~ 16:00
FITTING	1	3	4	5	2	6	7	8		

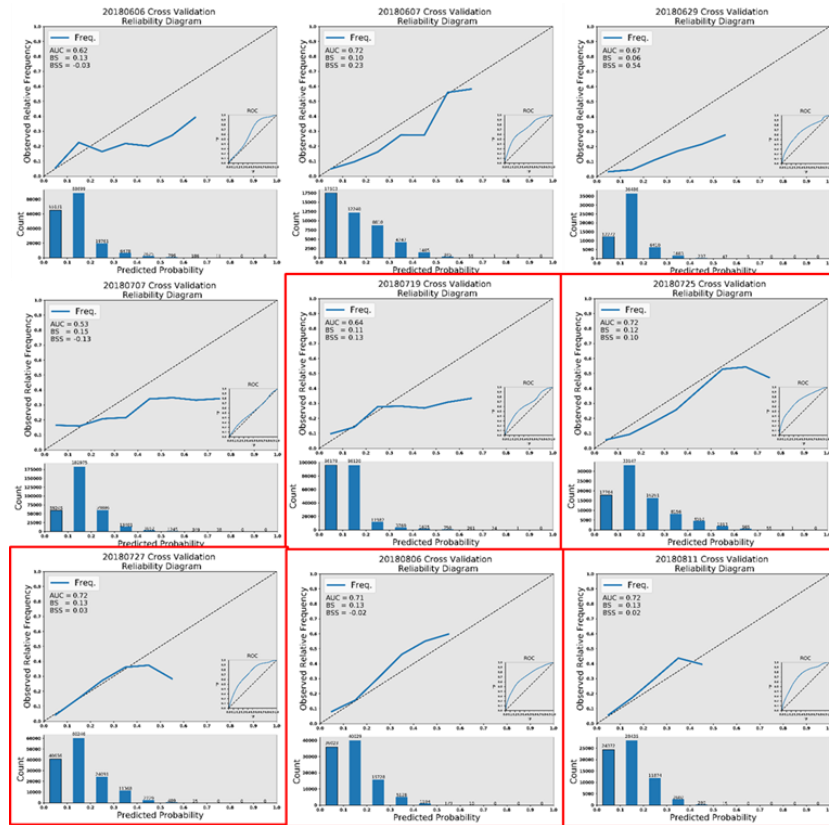


圖 1 2018 年 9 個午後對流個案交叉驗證可靠度、ROC 圖

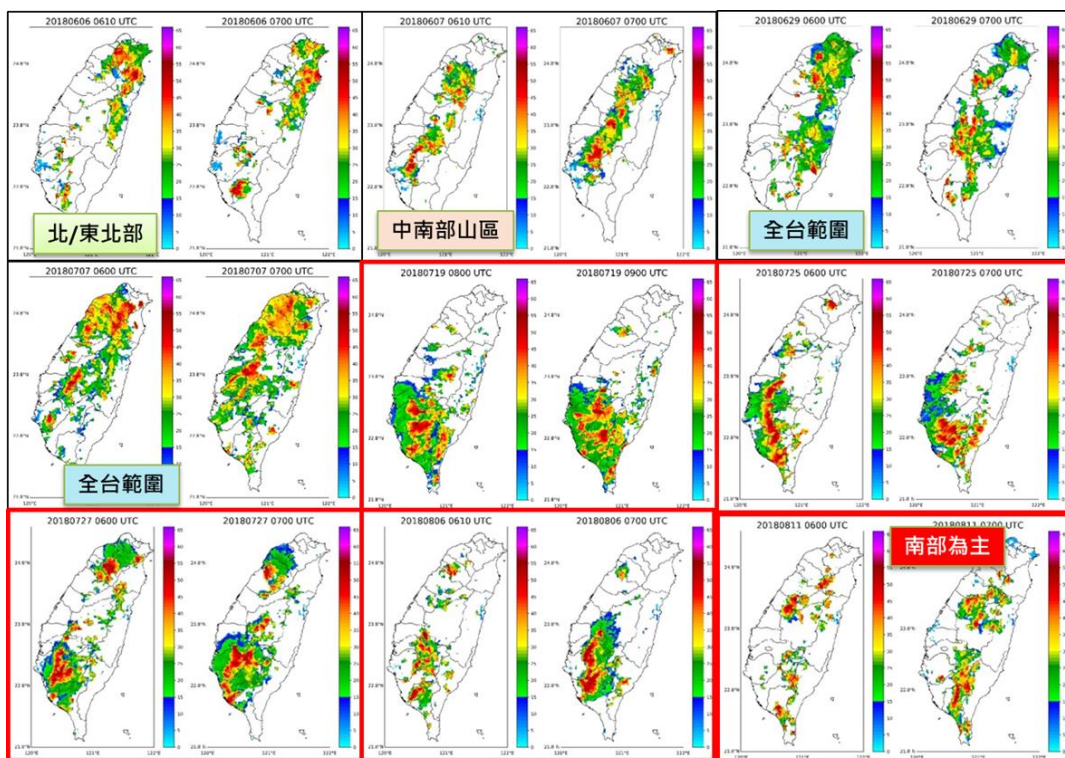


圖 2 2018 年 9 個午後對流個案雷達回波概況圖

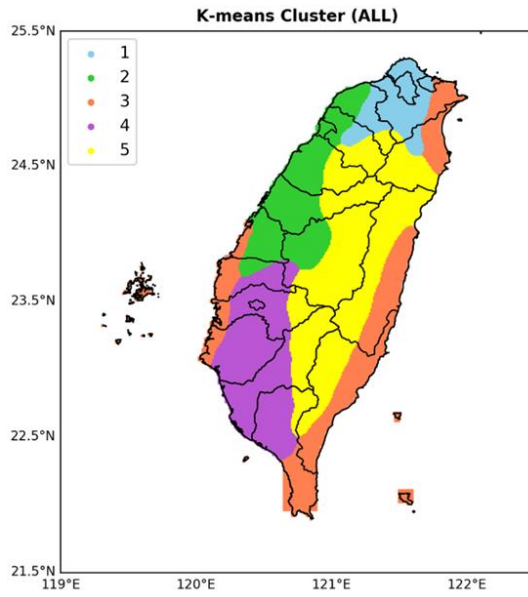
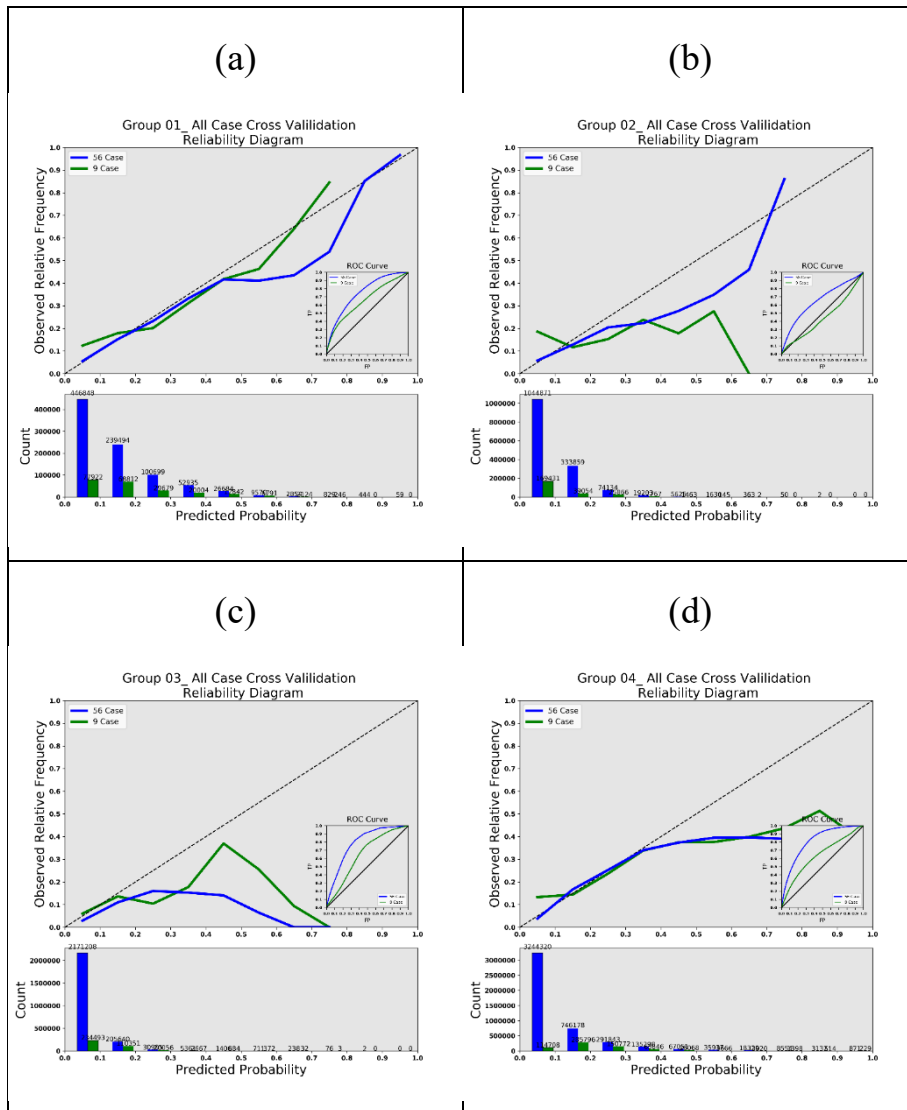


圖 3 K-means 分六群合併山區平滑後結果



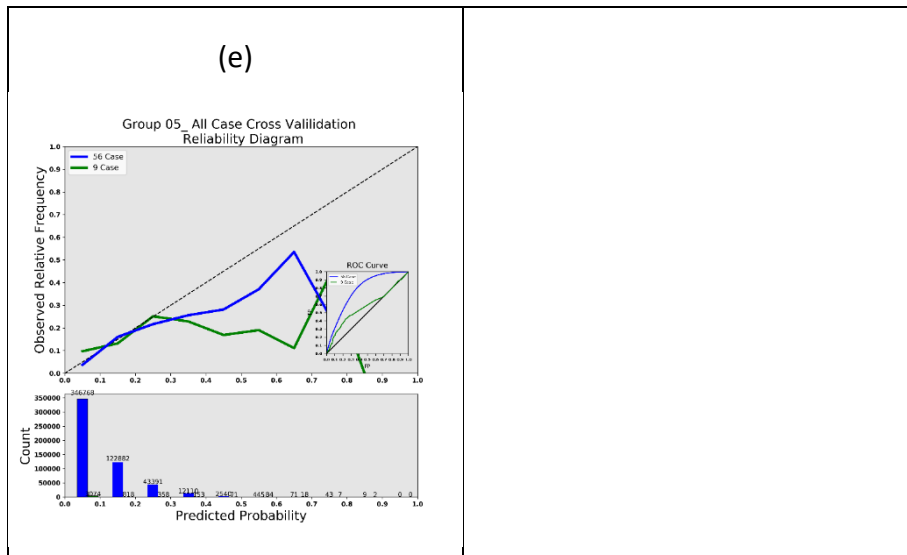


圖 4 (a)群 1(b)群 2(c)群 3(d)群 4(e)群 5 2018 年弱綜觀午後對流個案分群建模交叉驗證結果。
綠線:9 個案；藍線 56 個案

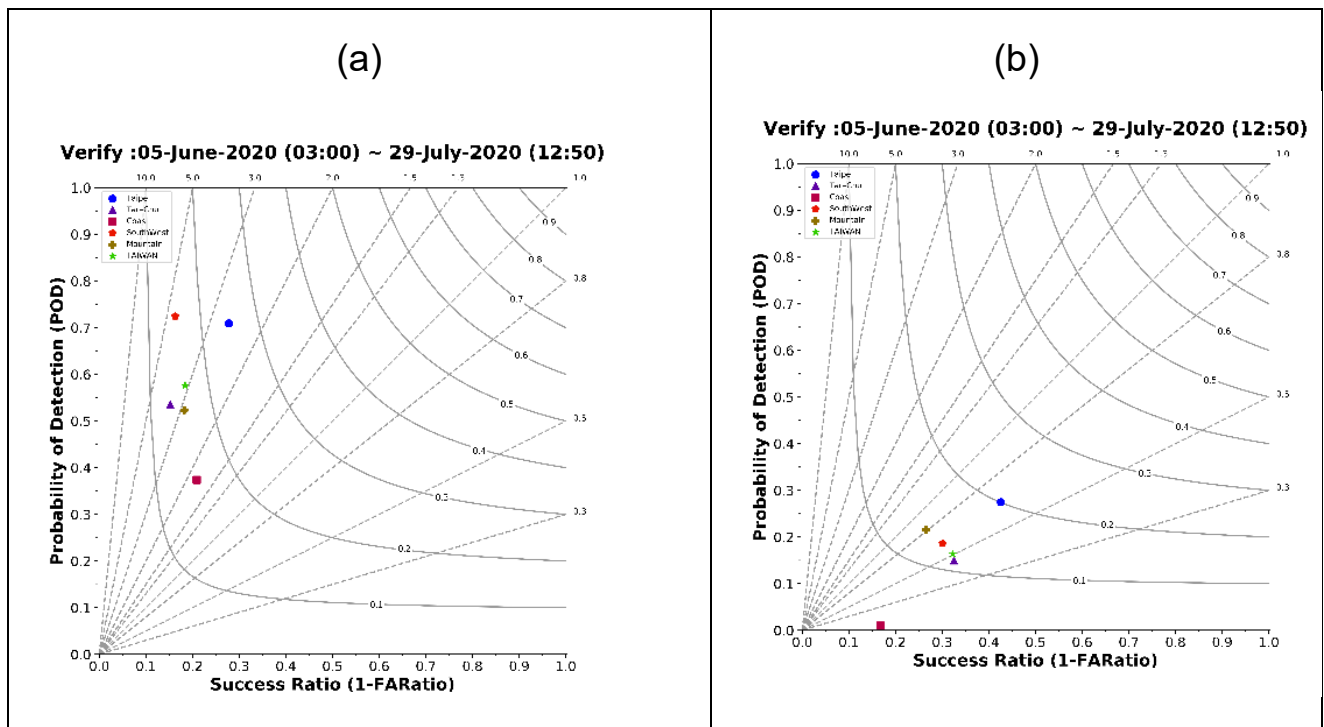


圖 4 (a)TANC 可能性預報(b)複邏輯斯迴歸機率預報

2020 年弱綜觀午後對流個案分群 Performance Diagram。

圖中的曲線是 TS 的等值線，而左下到右上的直虛線是 BIAS 的等值線。

Developing Probability Nowcasts of Convective Storm Initiation

Chih-Yung Feng¹, Yi-Han Chen¹, Han-Fang Lin¹, Yun-Tsai Lin², Wen-Ho Wang², Chia-Jung Wu², Yu-Ching Liu²

Manysplendid InfoTech, Co. Ltd.¹

Meteorological Satellite Center, Central Weather Bureau²

Abstract

To estimate the afternoon thunderstorm (radar echo > 35 dBz) initiation under weak synoptic-scale forcing, this study classified the Taiwan land grids into 5 categories by k-means clustering method and established multiple logistic regression for each category with predictors, including the starting position of convection, radar echo extrapolation, convective available potential energy, convective inhibition, average humidity, radar echo frequency, radar echo change trend, and surface divergence. Initiation probability of each grid (resolution 0.01 degree) for the next 60 minutes is provided every 10 minutes. Verification of cases in 2020 shows that it is not only comparable to Central Weather Bureau Taiwan AutoNowCaster (TANC), but also can further improve the forecast performance by including numerical model outputs, satellite products and high-resolution ground truth fields as predictors in the future.

Keyword: nowcasts, weak synoptic-scale forcing, probability forecast, afternoon thunderstorm, logistic regression