

貝氏模型平均法與多變量類比法應用於臺灣測站第3週均溫預報之校驗評估

林涵芳¹ 陳茂甄¹ 馮智勇¹ 羅存文² 張惠玲²
多采科技有限公司¹ 中央氣象局氣象科技研究中心²

摘要

現代的天氣預報常藉由數值系集模式預報描述大氣環境可能變化，然而系集預報卻普遍存有偏差或是離散度不佳而無法涵蓋真實事件的不確定性，需要藉由統計後處理方式提升預報能力。為開發臺灣測站第3週均溫預報指引，本研究使用EMC GEFS SubX模式溫度與遙相關指標層場資料，分別採用貝氏模型平均法(Bayesian Model Average, BMA)與多變量類比法(Analog Ensemble Post-processing, AP)建置均溫預報模型，並採1999~2019年遮蔽年實驗進行交叉驗證以評估兩方法機率預報表現。校驗結果顯示，BMA不論是在冬半年或夏半年皆可優於氣候值預報而具預報技術，而AP預報技術則於夏半年略不如氣候值預報。推測夏半年可能受梅雨與颱風系統影響，SubX模式預報臺灣測站第三週均溫較無穩定特性，而不適合應用AP統計後處理方法產製預報指引。

關鍵字：貝氏模型平均法、類比法、遙相關指標、機率預報

一、前言

從天氣到氣候的多重時間尺度中，目前天氣預報時效在2週以內，短期氣候預測則是在月時間尺度，2週到60天以內的氣象預報是發展無縫隙預報中最困難的時間尺度。從短期氣候預測的角度來看，發展季內尺度預測墜於無縫隙氣象預報服務相當重要，不僅可以填補中長期天氣預報和季節預測之間的空隙，也是連結天氣預報與短期氣候預測的橋樑。此外，外界對於預報產品需求也從模式格點降尺度到臺灣測站，因應在地化需求將進一步發展至臺灣高解析格點與農作物專區。

然而，現代的天氣預報通常基於數值模式模擬結果進行系集預報，其系集預報可能存有偏差或是低估分散度而無法捕捉真實事件的不確定性，故需要藉由統計後處理的方法提升預報能力，以期能提供測站週均溫完整機率密度函數(Probability Density Function, PDF)，可利於相關單位決策人員掌握未來第3週可能之平均溫度。

因此，本研究參考109年度「天氣與短期氣候統計後處理技術之研發委外案期末報告(2019)」中使用之多變量類比法(Analog Ensemble Post-processing, AP)改善系集模式降雨預報的空間偏差經驗，評估結合Kernel dressing發展測站未來第3週均溫機率預報的可行性。另一方面，本研究亦一

併評估貝式模型平均法(Bayesian Model Average, BMA)預報測站第3週均溫之表現。

二、資料說明與研究方法

(一) 資料說明

本研究模式預報使用EMC GEFS SubX模式之歷史預報資料(1999-2019年，不含2017年)，其預報頻率為每週1次(週三)，預報時間尺度為1日，預報至未來35天，水平與垂直解析度皆為 $1^\circ \times 1^\circ$ ，其中，1999-2016共有11組成員(含1組控制成員)，2018-2019共有21組成員(含1組控制成員)。測站資料採用25個人工站(表1)對應預報第3週之溫度資料。

表 1、臺灣 25 人工站碼、站名一覽表

站碼	站名	站碼	站名	站碼	站名			
1	466900	淡水	11	467350	澎湖	21	467610	成功
2	466910	鞍部	12	467410	台南	22	467620	蘭嶼
3	466920	台北	13	467440	高雄	23	467650	日月潭
4	466930	竹子湖	14	467480	嘉義	24	467660	台東
5	466940	基隆	15	467490	台中	25	467770	梧棲
6	466950	彭佳嶼	16	467530	阿里山			
7	466990	花蓮	17	467540	大武			
8	467060	蘇澳	18	467550	玉山			
9	467080	宜蘭	19	467571	新竹			
10	467300	東吉島	20	467590	恆春			

(二) 貝氏模型平均法(Bayesian Model Average, BMA)

BMA 該方法延伸自貝氏定理，藉由訓練期將機率預報命中觀測發生率最大化的演算法來調整各模式權重與條件機率分布參數。BMA 線性組合條件機率分布如式(1)所示，等號左側代表物理量 y 在已知 f_1 到 f_k 共 K 個預報結果時的發生機率分布，等號右側的 $p(y | f_k)$ 代表已知成員 k 預報值 f_k 時發生 y 觀測值的條件機率分布， $p(f_k | y^T)$ 則是在訓練資料 y^T 之中 f_k 表現為成員中最佳者的後驗機率，兩者相乘後加總，就能得到等號左側的物理量 y 發生機率密度函數估計。

$$p(y | f_1, \dots, f_k) = \sum_{k=1}^K p(y | f_k) p(f_k | y^T) \quad (1)$$

Raftery et al. (2005) 將條件機率以特定機率密度函數表示，並因後驗機率滿足 $\sum_{k=1}^K p(f_k | y^T) = 1$ 而視為權重係數 w_k ，則可藉由訓練期資料求得描述該機率密度函數所需的參數與權重。若使用常態分布 $g(y | f_k, \sigma_k)$ 表示該條件機率分布，即代表 y 的發生 機率滿足以 f_k 為中心的常態分布，可改寫上式如下，待估參數為 K 個預報成員的權重係數 w_k ，以及描述各個常態分布所需的參數標準差 σ_k ，可藉由最大化似然函數(Maximum likelihood, Fisher 1922)方式採 Expectation Maximization algorithm (EM 演算法) (Dempster et al., 1977; McLachlan and Krishnan, 1997) 求解待估參數。

$$p(y | f_1, \dots, f_k) = \sum_{k=1}^K w_k g(y | f_k, \sigma_k) \quad (2)$$

本研究之 BMA 使用觀測氣候值、模式溫度預報內插至觀測再與觀測溫度單迴歸與 GH, CuiXP, SCS 氣候指標各自與觀測溫度單迴歸(如圖 1)。建模策略採單一測站建模，並納入預報日前後兩週納入訓練。

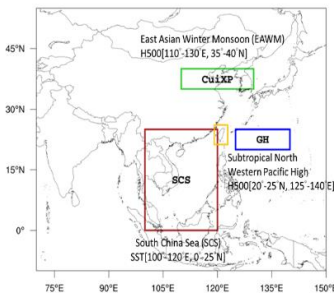


圖 1、BMA 溫度模型建置使用之遙相關指標(紅色範圍為 SCS、綠色範圍為 CuiXP、藍色範圍為 GH、黃色範圍為臺灣周圍 T2M)

(三) 多變量類比法(Analog Ensemble Post-processing, AP)

為說明多變數統計後處理的一般框架，定義氣象變數 $i \in \{1, \dots, I\}$ ，位置(如格點、測站) $j \in \{1, \dots, J\}$ 及預報

領先時間 $k \in \{1, \dots, K\}$ ， $l = (i, j, k)$ 及 $l^* = (i, j)$ 為對應變數、位置或預報領先時間的維度， $L = I \times J \times K$ 、 $L^* = I \times J$ ，系集成員個數為 M ，統計後處理挑選的成員數 N ，原始系集成員 $x = \{(x_1^1, \dots, x_M^1), \dots, (x_1^L, \dots, x_M^L)\}$ ，處理流程如下：

1. 建置相似資料集

Schefzik(2016)提出利用相似度準則挑選歷史上之觀測資料為相似資料集的做法。假設過去的時間點 t_d ， $d \in \{1, \dots, D\}$ ，則定義相似度準則：

$$\Delta^{t_d}(x^t, x^{t_d}) = \sqrt{\frac{1}{L^*} \sum_{l^*=1}^{L^*} (\bar{x}^{l^*,t} - \bar{x}^{l^*,t_d})^2 + \frac{1}{L^*} \sum_{l^*=1}^{L^*} (s^{l^*,t} - s^{l^*,t_d})^2} \quad (3)$$

$$\bar{x}^{l^*,t} = \frac{1}{M} \sum_{m=1}^M x_m^{l^*,t} \quad (4)$$

$$s^{l^*,t} = \sqrt{\frac{1}{M-1} \sum_{m=1}^M (x_m^{l^*,t} - \bar{x}^{l^*,t})^2} \quad (5)$$

其中， $\bar{x}^{l^*,t}$ 為格點系集平均， $s^{l^*,t}$ 為格點系集離散度。

評估時間點 t (目標日) 的系集與歷史系集的相似度，以格點系集平均誤差作為一階相似度，並以格點系集成員離散度作為二階相似度進行搜尋，即計算 $\Delta^{t_d}(x^t, x^{t_d})$ ，數值越小表示越相似。本研究比對臺灣周圍空間範圍溫度(如圖 1 黃色範圍)後，取最相似之前 20 組歷史預報對應之觀測作為相似系集成員。

2. 利用核函數綴飾法建置機率密度函數

核函數綴飾法可提供原始系集成員(動力性成員)額外成員，而這些後製成員有別於原始系集成員通常稱其為統計性成員，故因運算資源有限等因素無法大量產生動力性成員，將可使用核函數綴飾法去增加統計性成員，並藉此將原來離散分佈的系集成員轉換成連續型的機率分佈以進行機率預報。

核函數綴飾法基本概念為使用某特定機率分佈擬合系集成員，稱之為核函數(kernel)，再將這些核函數累加調整為一混合機率分佈，形式如下

$$p(y, x, \theta) = \frac{1}{d\sigma} \sum_i K \left[\frac{y - ax_i - \omega}{\sigma} \right] \quad (6)$$

其中， $x = [x_1, \dots, x_d]$ 為原始系集成員， d 為系集成員個數， y 為觀測值或分析場，每個核函數位置中心為 $ax_i + \omega$ ， a 、 ω 與 $\sigma (> 0)$ 分別是系集成員 x_i 的 scale factor、offset factor 與 bandwidth factor。

由式(6)可知，核函數擬態需定義 bandwidth factor，故本計畫以經驗公式作為 bandwidth factor 之依據，

$$bw = \left(\frac{4\sigma^5}{3n}\right)^{\frac{1}{5}} \approx 1.06\sigma n \quad (7)$$

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (8)$$

其中， σ 為測站成員標準差， x_i 為測站成員， n 為成員數量。

三、評估系集機率預報能力

(一) 校驗方法

本研究校驗方法利用逐年交叉驗證(遮蔽年預報)，並根據 RPSS(Ranked Probability Skill Score)、CRPSS(Continuous Ranked Probability Skill Score, Hersbach, 2000)、Reliability diagram、CS(Calibration Score, Krzysztofowicz, 1999)與 IS(Informativeness Score, Krzysztofowicz, 2008)等機率預報技術評估指標進行分析。本研究採用之三分類門檻為 1989-2019 (移除校驗年)的觀測資料計算之 33、67%門檻值。

(二) 預報能力評估

根據圖 2 與圖 4 測站 RPSS 與 CRPSS 結果顯示，BMA 相對於氣候預報而言，具有良好的預報技術，然而，AP 於夏半年之預報技術較氣候預報差，可能是由於 AP 預報直接使用模式空間範圍溫度，若受梅雨季與颱風影響，模式較不易掌握溫度，導致模式溫度預報空間分布與真實測站溫度分布不同時，進而增加 AP 方法應用上之困難。此外，亦可能是 AP 未使用氣候值作為參數，導致夏半年預報技術略低於氣候預報。

另一方面，根據測站 IS 差異結果顯示(如圖 3)，不論是 BMA 或 AP，兩者皆具有潛在的經濟價值(IS 值皆大於 0.8 以上)，其中以 BMA 於所有季節與夏半年之潛在價值略具優勢。由圖 5 測站 CS 差異與可靠度圖結果顯示，以機率預報校正技術而言，不同預報方法皆具有良好的預報校正技術，且 AP 與 BMA 不相上下。以 Reliability diagram 之可信度而言(如圖 6)，BMA 於 Below、Normal、Above 皆具有良好的可信度，而 AP 則容易於低機率時低估、高機率時高估之情形。

四、結論與建議

綜合上述結果，BMA 於各個機率評估指標皆具有相當穩定的預報技術，而 AP 預報受限於方法本身之限制，導致 AP 預報技術於夏半年略不如氣候值預報。推測夏半年可能受梅雨與颱風系統影響，SubX 模式預報臺灣測站第 3 週均溫較無穩定特性，而不適合應用 AP 統計後處理方法產製預報指引。

五、參考文獻

1. 交通部中央氣象局，2020：氣象模式統計後處理技術之研發期末報告，第陸章，89-112
2. Fisher, R. A., 1922: On the mathematical foundations of theoretical statistics. *Philos. Trans. Roy. Soc. London*, 222A, 309–368.
3. Dempster, A. P., N. M. Laird, and D. B. Rubin, 1977: Maximumlikelihood from incomplete data via the EM algorithm. *J.Roy. Stat. Soc.*, 39B, 1–39.
4. Kelly, K.S. and R. Krzysztofowicz, 1995: Bayesian revision of an arbitrary prior density., *Proceedings of the Section on Bayesian Statistical Science*, American Statistical Association, 50-53
5. McLachlan, G. J., and T. Krishnan, 1997: *The EM Algorithm and Extensions*. Wiley, 274 pp.
6. Krzysztofowicz, R., and A. A. Sigrest, 1999: Calibration of probabilistic quantitative precipitation forecasts. *Wea. Forecasting*, 14, 427–442.
7. Hersbach, H., 2000: Decomposition of the continuous ranked probability score for ensemble prediction systems. *Wea. Forecasting*, 15, 559–570.
8. Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Polakowski, 2005: “Using Bayesian model averaging to calibrate forecast ensembles”. *Mon. Wea. Rev.*, 133, 1155–1174.
9. Krzysztofowicz, R., and W. Evans, 2008: Probabilistic Forecasts from the National Digital Forecast Database. *Weather and Forecasting*, 4(27): 270-289.
10. Krzysztofowicz, R., 2010: Decision criteria, data fusion and prediction calibration: a Bayesian approach. *Hydrological Sciences Journal*, 55(6): 1033-1050

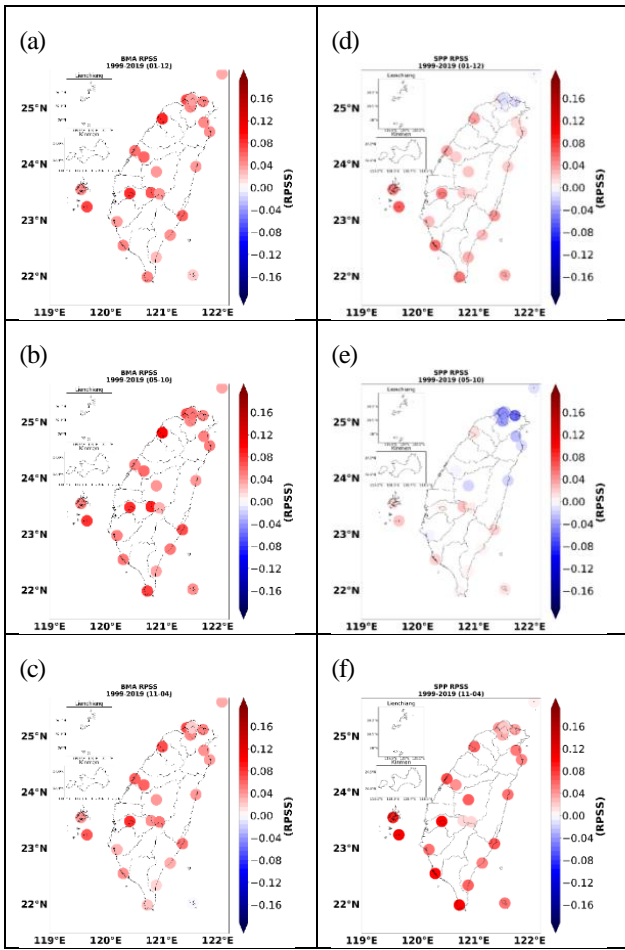


圖 2、測站 RPSS。左為 BMA、右為 AP，由上至下分別為校驗期，所有季節、夏半年(05-10月)與冬半年(11-04月)。(完美技術得分 RPSS 為 1，無預報技術為 0)

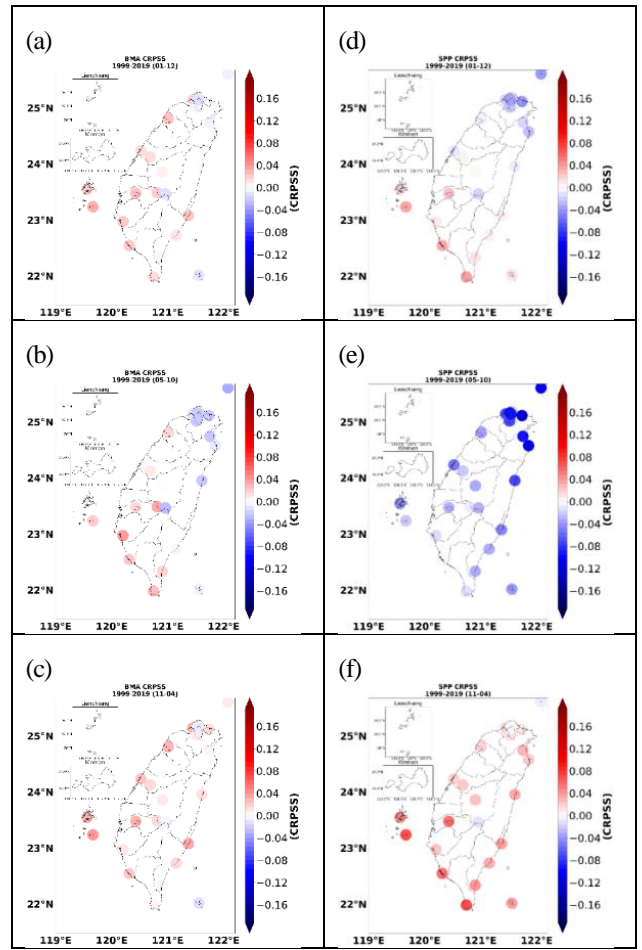


圖 4、圖說同圖 2，但為測站 CRPSS。(CRPSS 大於 0 表示相較於氣候值，模式具有預報能力)

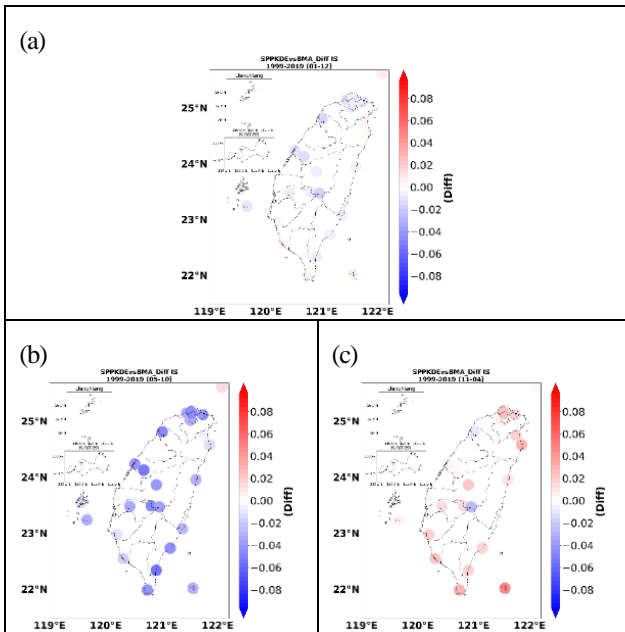


圖 3、BMA 與 AP 之測站 IS 差異。(a)所有季節、(b)夏半年、(c)冬半年。越藍表示 BMA 潛在經濟價值越高。

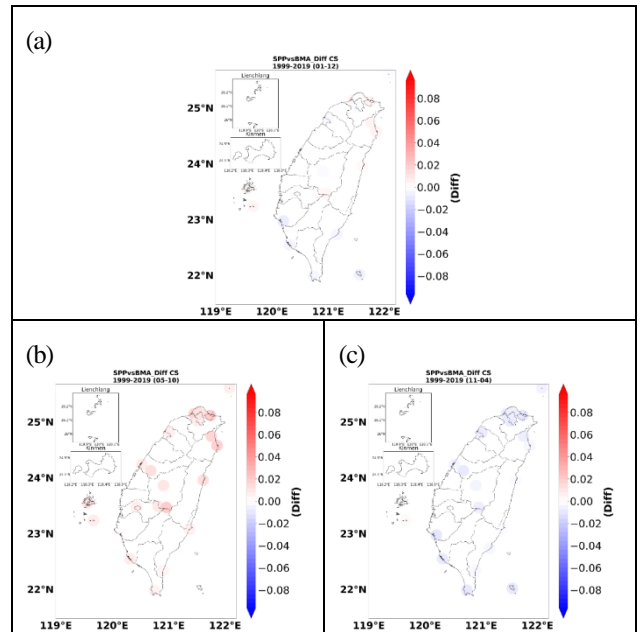


圖 5、圖說同圖 3，但為 BMA 與 AP 之測站 CS 差異。越紅表示 BMA 具有良好的校正性質(Well-Calibrated)。

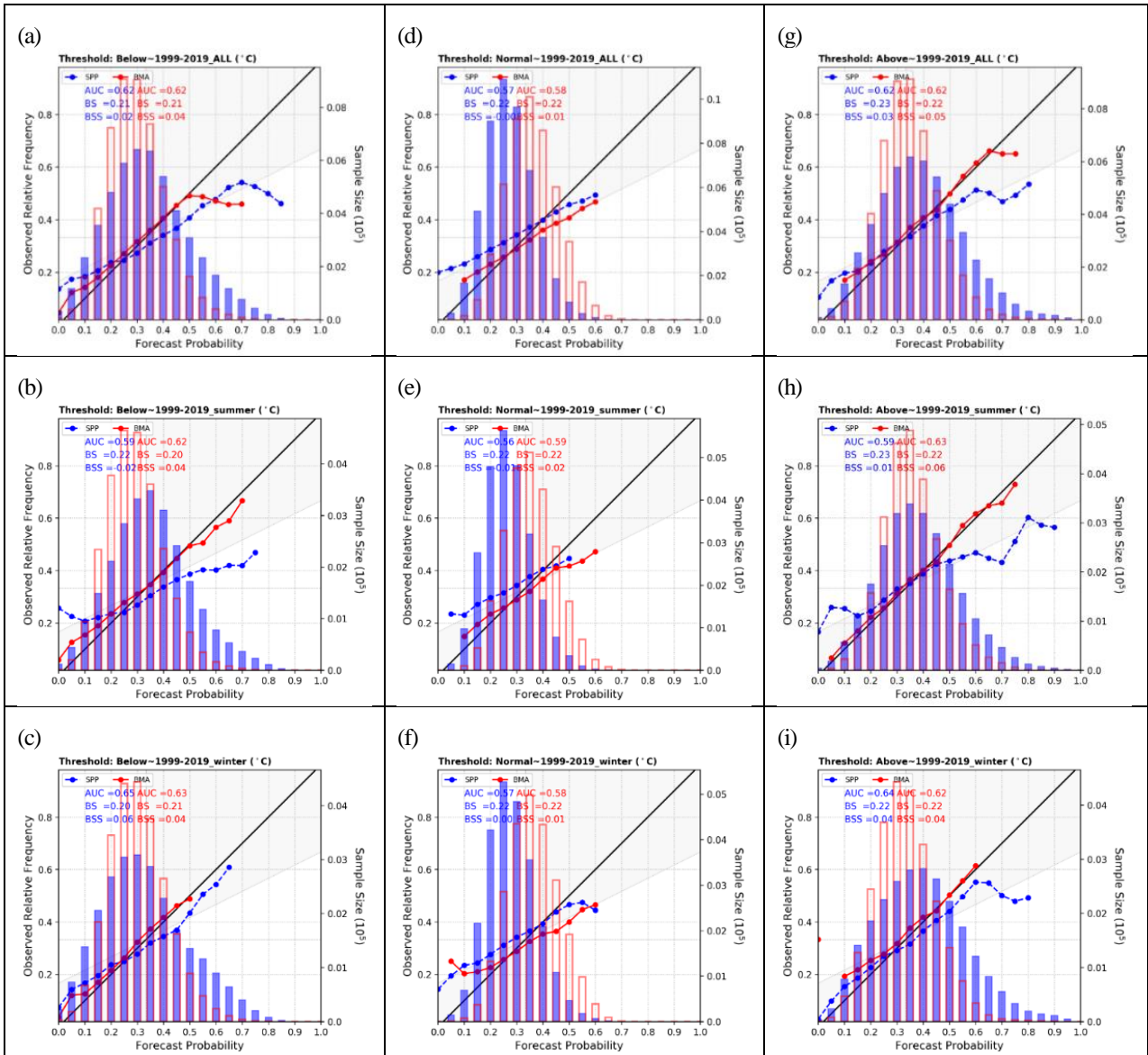


圖 6、BMA(紅)與 AP(藍)之三分類機率 Reliability diagram。由左至右分別為 Below、Normal、Above，上至下分別為所有季節、夏半年、冬半年。橫軸為機率區間，左縱軸為觀測發生頻率(對應信賴曲線)，右縱軸為預報樣本數(對應樣本長條圖)。

Evaluation of Bayesian Model Average and Analog Ensemble Post-Processing Forecasts for Week 3 Temperature on Taiwan Stations

Lin H.-F¹, Chen Y.-C¹, Feng C.-Y¹, Lo T.-W², Chang H.-L.²

¹Manysplendid Infotech Ltd., Taiwan

²Research and Development Center, Central Weather Bureau, Taiwan

Abstract

Weather forecasts are usually performed based on numerical ensemble model outputs to capture the possible perturbations of atmospheric environment. Numerical ensemble model, however, may not depict the uncertainty of real weather events due to the existence of systematic biases or poor model output spread. Therefore, statistical post-processing methods are always implemented to improve the model capability of extended range forecasting. To provide the forecast guidance for mean temperature during week 3 on Taiwan stations, this study uses 1999-2019 EMC GEFS SubX model temperature and teleconnection-related variables output data for developing two kinds of statistical downscaling forecast, namely the Bayesian Model Average (BMA) and the Analog Ensemble Post-processing (AP), and conducts a leave-one-year-out cross validation experiment to evaluate the probability forecast performances. Results show that BMA forecast is better skillful than the climate forecast for the whole year, while the skill of AP forecast is slightly less than the climate one in the summer half year. Consider the deficiency of resolution and mechanism, SubX model may suffer from predicting Mei-yu and typhoon events over Taiwan with consistency and therefore limits the performance of AP.

Key words: Bayesian Model Average, Analog Ensemble, teleconnection index, probabilistic forecasts