

# 貝氏模型平均應用於臺灣地區溫度機率預報

張語軒、張庭槐、吳蕙如

氣象資訊中心

中央氣象局

## 摘要

系集預報的目的是希望盡可能掌握天氣發展的不確定性，尋找天氣未來狀況可能的分散範圍，若能參考實際天氣的機率分布，則能更精確地描述此不確定性。

本研究首先利用統計方法貝氏模型平均（Bayesian model averaging，簡稱BMA）將多組模式的預報進行後製處理，建立符合天氣真實狀況的預測機率密度函數（probability density function，簡稱PDF）。

其後將BMA應用在臺灣地區測站地面溫度的預報上，經過校驗發現BMA的預測PDF在預報表現上很理想，儘管多模式預報成員的預報值離散度不夠，但BMA還是能有效掌握溫度變化相當程度的不確定性。

關鍵字：貝氏模型平均、機率密度函數、系集預報

## 一、前言

我們永遠無法得知未來真正的天氣狀態，天氣的演變充滿著許多不確定的因素，要得到一個完全準確的決定性預報，不僅不容易達到，在運用上亦是相當冒險的，因此期望將數值模式預報的結果藉由系集的方法盡可能呈現未來天氣之不確定性。然而，如何透過預報的結果將不確定性表達出來，是一個困難卻又重要的問題。若能從數值預報系統中建構出預測值的預測機率密度函數（probability density function，簡稱

PDF），是探究不確定性很好的一個方式，透過PDF可以很清楚地了解到未來觀測值所有可能發生的範圍及其機率，進而能呈現出更精確的氣象預報。

傳統的系集預報，其發展的核心概念在於假設每一個系集預報成員以及觀測值皆來自相同的機率分布。此假設隱含著每一個模式都有相同的準確度，且模式的分散範圍可有效的掌握觀測值的不確定性，進而以此概念建立不同的系集預報產品。例如，要利用多個模式預報成員來建立決定性預報，大多使用系集預報成員的平均值；而利用多組

模式預報成員建立預測機率分布，則採用 Anderson於1996年提出的binned probability ensemble (簡稱BPE) 來進行。

BPE方法的預測機率分布是：若預報系統中有  $K$  個預報成員，將這些成員產出的  $K$  個預報值依其大小，在  $-\infty$  到  $+\infty$  的數線上從小到大依序排列，可構成  $(K+1)$ 個區間，則觀測值會有相同的機率落在這  $(K+1)$ 個區間中，即落在每一個區間中的機率皆為  $1/(K+1)$ 。

但我們認為上述的傳統方法仍有許多不足的地方，例如：透過校驗會發現利用BPE得到的機率分布往往與實際狀況差異很大，其主要原因是預報成員通常都會有離散程度不夠的現象；透過校驗也可發現並非每一個預報成員皆有相同的準確度。而更大的問題是傳統方法無法建立起完整的預測PDF，這正是本文最主要是想解決的問題。

本文採用Hoeting等人 (Hoeting et al., 1999) 提出的貝氏模型平均法 (Bayesian model averaging, 簡稱BMA)，針對多組模式預報進行後製處理，建構出一個有效的預測PDF，來改善傳統方法不足的地方。

BMA已是被廣泛應用於經濟學、生物學、公共衛生學等領域的統計方法，近期也漸漸運用在氣象領域而有許多文獻發表。Reftery et al. (2005) 採用MM5 (The PSU / NCAR meso-scale model) 透過不同初始條件和側邊界條

件產生的預報資料，將BMA應用在美國西南部地面氣溫48小時的預報，經過校驗顯示BMA建立出來的預測機率分布比傳統方法表現更好，其應用在海平面氣壓也是類似的結果。Sloughter et al. (2007、2010) 亦將BMA分別應用在降雨量及風速的預報上，表現都很不錯。而在統計自由軟體R，BMA已有套件可供使用 (Fraleley et al. 2007)，足見BMA的應用已越來越廣被接受。

本篇文章首先將對BMA進行簡短的介紹，接著將BMA應用於本局區域模式在臺灣地區近地面溫度的24小時預報資料，建立溫度的預測PDF並進行校驗，最後介紹BMA產製的PDF在實際預報上的應用。

## 二、BMA介紹

### (一) 模型介紹

BMA模型是利用在得知預報成員的情況下，預測出對於觀測值的條件機率密度函數 (conditional probability density function, 簡稱 conditional PDF)。

若令  $y$  為我們想要預測的天氣變數，而  $f_1, f_2, \dots, f_K$  為已知的  $K$  個預報成員的預報值，則BMA模型下的預測PDF表示為：

$$p(y | f_1, f_2, \dots, f_K) = \sum_{k=1}^K w_k g_k(y | f_k), \quad (1)$$

其中  $w_k$  為第  $k$  個成員的後驗機率 (posteriori probability)，代表在  $K$  個預報成員中，

第  $k$  個成員預報最佳的機率，其滿足  $\sum_{k=1}^K w_k = 1$  的條件。而  $g_k(y|f_k)$  為條件PDF，代表在第  $k$  個成員預報最佳且其預報值為  $f_k$  的條件下  $y$  的PDF。

BMA的預測PDF可視為多個PDF的加權平均，以圖1進行說明。圖中表示共有3個預報成員，每一條虛線分別是每一個預報成員對應到的  $g_k(y|f_k)$ 。若以每個成員的後驗機率  $w_k$  為權重進行加權平均，則得到圖中的粗實線，此即為BMA模型下的預測PDF，也就是公式(1)中的  $p(y|f_1, f_2, \dots, f_K)$ 。

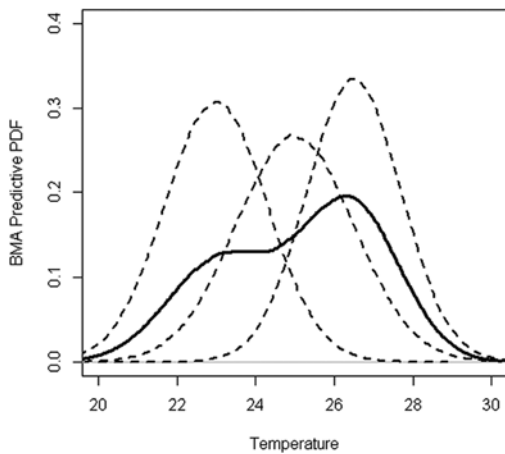


圖1. 以溫度場為例，採用BMA方法下的預測PDF示意圖。圖中的虛線代表每一個模式對應到的條件PDF，將三條虛線做加權平均後，得到的實線即為觀測值的預測PDF。

Figure 1. Schematic diagram shows the BMA predictive PDF for temperature forecast. The dotted lines are the conditional PDFs for each model. The solid line is the BMA predictive PDF, which is a weighted average of the conditional PDFs.

若是溫度、氣壓等變數，其數據分布相當對稱，會非常接近常態分布 (normal distribution)，因此可假設  $g_k(y|f_k)$  為常態分布的PDF，期望值為  $a_k + b_k f_k$ 、標準差為  $\sigma$ ，其中  $a_k$  和  $b_k$  是對預報值進行偏差調整時採用的參數。若用機率符號表示則是

$$y|f_k \sim N(a_k + b_k f_k, \sigma^2)。$$

在利用BMA求出  $y$  的預測PDF (公式1) 之後， $y$  的期望值為

$$E(y|f_1, f_2, \dots, f_K) = \sum_{k=1}^K w_k (a_k + b_k f_k) ， \quad (2)$$

其可用來做為  $y$  的決定性預報。

由於本文探討的是溫度預報，因此採用常態分布假設下的BMA。若是針對雨量或是風速這一類非對稱的數據，可改用其它函數，例如gamma分布，來進行逼近，而關於gamma分布假設下的BMA模型，可參考Sloughter et al. (2007、2010) 文章中詳盡的描述。

## (二) 參數估計

在常態分布的假設下，BMA模型有  $a_k$ 、 $b_k$ 、 $w_k$  ( $k=1, \dots, K$ ) 及  $\sigma$  這些參數，我們將預報發佈的前  $N$  天視為訓練期，利用訓練期的數據來估計參數。訓練期的長短是可變的，關於  $N$  的挑選問題將於第四節再做說明。在設定訓練期  $N$  之後，參數的估計方法本文依照 Reftery et al. (2005) 的做法，將使用統計方法如最小平方法 (least squares method)、最大概似估計法 (maximum likelihood estimation) 進行估計，以下說明估計步

驟。

首先，我們利用簡單線性迴歸模型對模式預報進行偏差調整。在  $s$  地點及  $t$  時間下，假設觀測值  $y_{st}$  跟第  $k$  個模式預報值  $f_{kst}$  的關係可表示成下列式子：

$$y_{st} = a_k + b_k f_{kst} + \varepsilon_{ks}$$

其中  $\varepsilon_{ks}$  為誤差。再利用訓練期中的資料來估計  $a_k$  及  $b_k$ ，估計的方法採用最小平方法，得到

$$b_k = \frac{\sum_{s,t} (f_{kst} - \bar{f}_k)(y_{st} - \bar{y})}{\sum_{s,t} (f_{kst} - \bar{f}_k)^2},$$

$$a_k = \bar{y} - b_k \bar{f}_k,$$

其中

$$\bar{y} = \frac{1}{n} \sum_{s,t} y_{st}, \quad \bar{f}_k = \frac{1}{n} \sum_{s,t} f_{kst},$$

$n$  為訓練期中所有數據的個數。

接著，對於  $w_k$  及  $\sigma$  的估計則使用最大概似估計法。根據公式1我們可得到對數概似函數 (log-likelihood function) 為

$$l(w_1, \dots, w_K) = \sum_{s,t} \log \left( \sum_{k=1}^K w_k g_i(y_{st} | f_{kst}) \right),$$

再透過EM (Expectation-Maximization) 演算法來尋找最大概似估計值。

EM演算法的做法如下：

設一個潛在變數  $z_{kst}$ ，當  $z_{kst}$  等於1代表在  $s$  地點及  $t$  時間下  $f_{kst}$  為最佳預報值，反之則  $z_{kst}$  為0。對於欲估計的參數  $w_k$  及  $\sigma$ ，先給定一個初始猜測值，然後透過重複地疊代運算直到參數收斂。在第  $j$  次的疊代運算中，

$$\hat{z}_{kst}^{(j)} = \frac{w_k g(y_{st} | f_{kst}, \sigma^{(j-1)})}{\sum_{i=1}^K w_i g(y_{st} | f_{ist}, \sigma^{(j-1)})},$$

其中  $g(y_{st} | f_{kst}, \sigma^{(j-1)})$  為常態分布的PDF，其期望值為  $a_k + b_k f_{kst}$ ，標準差為  $\sigma^{(j-1)}$  的。計算出  $\hat{z}_{kst}^{(j)}$  之後，再用  $\hat{z}_{kst}^{(j)}$  來計算  $w_k$  及  $\sigma$ ，得到

$$w_k^{(j)} = \frac{1}{n} \sum_{s,t} \hat{z}_{kst}^{(j)},$$

$$\sigma^{2(j)} = \frac{1}{n} \sum_{s,t} \sum_{k=1}^K \hat{z}_{kst}^{(j)} (y_{st} - f_{kst})^2.$$

以上即為本文中，BMA模型參數的完整估計流程。

### (三) 全區及分區BMA模型

在Reftery et al. (2005) 的文章中，對於同一個預報時間，各個測站點的BMA模型都是使用同一組參數 ( $a_k$ 、 $b_k$ 、 $w_k$  及  $\sigma$ )。但經過我們的比較，發現對於每一個地點分別去計算不同的BMA模型參數，會有更好的表現。即是將模型參數分成  $a_{ks}$ 、 $b_{ks}$ 、 $w_{ks}$  及  $\sigma_s$ ，下標  $s$  代表不同的地點，而每一個地點只使用訓練期中該點的資料來估計參數。在此，本文將這兩種方法分別稱為全區BMA模型以及分區BMA模型，之後將會比較這兩種BMA的表現。

## 三、校驗方法介紹

以下將介紹幾項校驗方法，對於每一個校驗方法，我們都會拿BMA與傳統方法進行比較。傳統方法是建立在所有預報成員與觀測值皆來自相同機率分布的假設下所發展出來，其給與每一個預報

成員相同的權重，且利用預報成員的分散範圍來預測觀測值的可能變異。

### (一) Rank Histogram及Probability Integral Transform

首先介紹的校驗方法，是利用圖形法來校驗預測機率分布。由於傳統方法建立出來的預測機率分布為離散型，而BMA方法建立出來的預測機率分布為連續型，因此兩者在校驗上使用的圖形法有些微的不同。

傳統方法使用BPE來建立預測機率分布，也就是假設觀測值會有相同的機率落在  $K$  個預報成員構成的  $(K+1)$  個區間中。而在Anderson (1996) 文章中提到BPE的檢驗，可藉由rank histogram (RH) 來觀察此機率分布是否準確。RH是藉由計算觀測值落在這  $(K+1)$  個區間的比例所畫出的直方圖，如圖2所示。圖中橫軸代表的是累積機率，因為觀測值落在每一個區間的機率皆假設為  $1/(K+1)$ ，所以每一個塊型的寬度皆為  $1/(K+1)$ ，而每一個塊型面積則分別代表觀測值落在這  $(K+1)$  個區間的比例。若觀測值有相同的機率落在這  $(K+1)$  個區間中，則得到的RH圖將會很平坦，完美情況下每一個塊型高度會很接近圖2中的虛線。

傳統方法的離散型預測機率分布可藉由RH進行校驗，若是針對BMA模型下的連續型預測機率分布，Gneiting et al. (2005) 採用與RH類似的probability integral transform (PIT) 直方圖進行校驗，PIT可視為RH的連續型，其做法如

下：

設  $p(x)$  代表BMA下的預測PDF， $P(x)$  為其對應的累積分布函數 (cumulative distribution function，簡稱CDF)， $P(x)$  與  $p(x)$  的關係為

$$P(x) = \int_{-\infty}^x p(y) dy。$$

而將所有觀測值  $x_o$  對應到的  $P(x_o)$  畫成直方圖，即為PIT直方圖，如圖2所示。

若我們的預測PDF很準確，則對應到

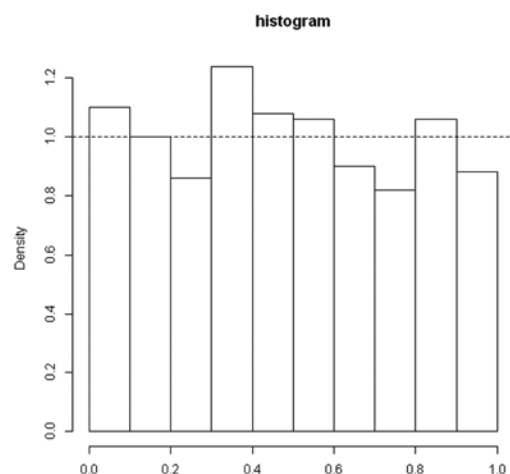


圖2. RH以及PIT示意圖。圖中橫軸為累積機率，每一個塊型寬度為觀測值落在該區間的理論機率值，而塊型面積則代表觀測值落在該區間的實際比例。虛線代表在完美情況下，直方圖的高度。

Figure 2. Schematic diagram of RH and PIT histogram. The horizontal axis is cumulative probability. The width of the rectangle is the theoretical probability for the interval to coverage observation, and the area of the rectangle is the actual frequency for the interval to coverage observation. The dotted line is the perfect height for the histogram.

的PIT直方圖會是很平坦的，其機率理論是因為對於任意隨機變數 $X$ ，所對應到的累積機率 $P(X)$ 會呈現均勻分布（uniform distribution）。此外，PIT直方圖中的塊形個數可任意設定，不同於RH的塊型個數固定為 $(K+1)$ 。

雖然PIT與RH兩個直方圖的計算方式不同，但兩者的機率概念其實是一樣的，差別只在於RH是用來處理離散型的機率分布，而PIT是用來處理連續型的機率分布。因此RH與PIT直方圖的解釋方法相同，兩者可以直接進行比較。

## (二) Continuous ranked probability score

Continuous ranked probability score (CRPS) (Hersbash, 2000) 是利用數值法來校驗預測CDF。

CRPS的計算方法是使用布賴爾評分 (Brier score) 對所有門檻值積分，CRPS的公式為

$$CRPS = \int_{-\infty}^{\infty} [P(x) - O(x)]^2 dx, \quad (3)$$

其中  $P(x)$  為預測CDF，而

$$O(x) = \begin{cases} 0, & x < x_o \\ 1, & x \geq x_o \end{cases},$$

$x_o$  為觀測值。CRPS可視為預測CDF準確性的得分，其值愈低代表預測CDF越準確。

使用CRPS來校驗BMA方法時，公式(3)中的  $P(x)$  是將BMA建立出來的預測PDF轉換為CDF後代入，如圖3中的上圖所示。當校驗傳統方法時， $P(x)$  則是使用成員預報值的經驗累積分布函

數 (empirical cumulative distribution function, 簡稱ECDF)，如圖3中的下圖所示，圖中每一個階梯上升的位置代表某一成員的預報值。由於傳統方法假設每一個預報成員與觀測值來自相同的機率分布，因此其預測CDF為其ECDF。

## (三) 預測區間涵覆比例及寬度

利用BMA方法下得到的預測PDF，可以計算出任何百分比的預測區間。例如取出PDF上5%和95%分位數可視為90%預測區間，接著可再計算觀測值落在90%預測區間的真實比例，看是否接近理想中的90%。由於BMA可以建立連續型的預測PDF，因此透過BMA可以得到任意百分比的預測區間，而離散型的BPE方

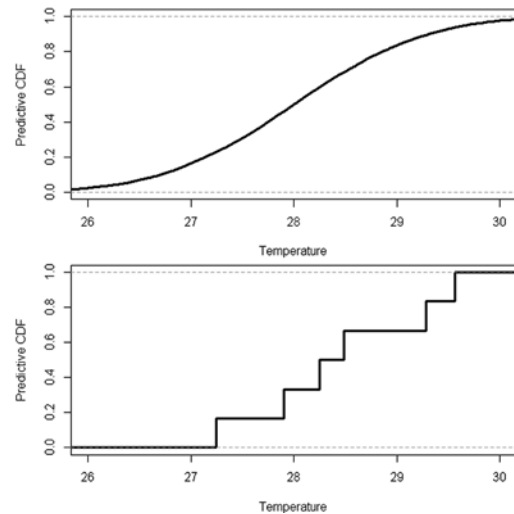


圖3. 預測CDF示意圖。上圖代表經由BMA得到的預測CDF，下圖代表傳統方法使用的預測CDF。

Figure 3. Schematic diagrams show the predictive CDF. The upper plot shows the BMA predictive CDF, and the lower plot shows the traditional predictive CDF.

法則不行。在6個模式成員的情況下，傳統做法中BPE假設觀測值有相同的機率會落在預報值所構成的7個區間內，因此可以取6個預報值中的最大值和最小值做為71.4%預測區間（ $5/7=71.4\%$ ）。除了校驗預測區間的涵覆比例，還可校驗預測區間的寬度。若有兩種預報其預測區間在涵覆比例上的表現差異不大時，通常會期望區間的寬度越窄越好，因為寬度越寬表示預報的不確定性也就越高。

#### (四) 平均絕對誤差

平均絕對誤差（mean absolute error，簡稱MAE）是用來校驗決定性預報與觀測值之間的誤差，不同於上述校驗方法主要的對象為預測機率分布。MAE的公式為

$$\frac{1}{n} \sum_{i=1}^n |f_{di} - y_i|,$$

其中  $f_d$  為決定性預報，而  $y$  為觀測值，其值越低越好。對於傳統方法，一般採用預報成員的平均值當成決定性預報，而BMA的決定性預報在此是採用預測機率分布的期望值（見公式2）。此外，BMA也可使用中位數來做為決定性預報。

## 四、實際分析與校驗

### (一) 資料來源

本文使用的資料為臺灣地區17個綜觀測站的地面溫度觀測值及本局區域模式內插至觀測點上的第24小時預報值，測站包含淡水、台北、竹子湖、基隆、松山、花蓮、澎湖、台南、高雄、

嘉義、台中、阿里山、玉山、新竹、蘭嶼、日月潭、台東，資料的時間為2010年10月20日到2011年6月7日，使用的模式資料為2組NFS（Non-hydrostatic Forecast System）、1組WRF（Weather Research & Forecast model）的預報，每組模式都分別包含解析度5公里及15公里的預報資訊，共有6群模式預報資料，將這些格點資料內插至測站後即是預報的成員。

### (二) BMA模型訓練期測試

在建立BMA模型時，須先設定模型訓練期  $N$ ，而關於  $N$  的挑選，Reftery et al.（2005）的研究中並未有客觀的訂定準則，我們利用測試各種不同訓練期下模型的表現來選定  $N$  值，測試時間為2010年12月22日到2011年6月7日。

首先觀察全區BMA模型的訓練期測試，測試的訓練天數共有10、15、20、25、30、35、40、45、50等9個，其結果如圖4所示，圖中包含校驗CRPS、MAE、66.7%和90%預測區間實際涵覆比例及66.7%和90%預測區間的寬度。從圖4(a)中，可以發現若以20天為訓練期CRPS的表現最好，而圖4(b)中則顯示20天和25天的訓練期有較佳的MAE（兩者差距微小，僅有 $5 \times 10^{-5}$ ）。圖4(c)和圖4(d)分別是描述66.7%和90%預測區間涵覆觀測值的比例，圖4(c)裡66.7%預測區間皆高於66.7%，其中以20天為訓練期的涵覆比例比理想值多了2.9%。但圖4(d)裡預測區間的涵覆比例都相當接近90%，

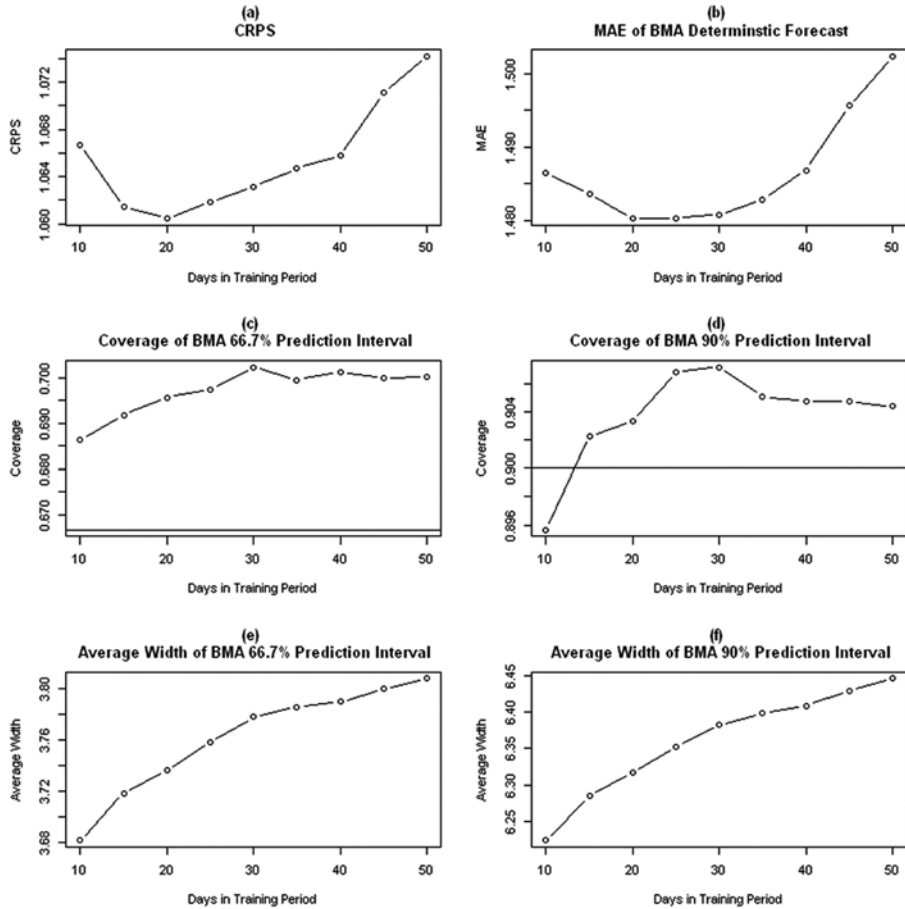


圖4. 全區BMA模型訓練期測試。(a) RPS，(b) MAE，(c) 66.7%預測區間的涵覆比例，(d) 90%預測區間的涵覆比例，(e) 66.7%預測區間的平均寬度，(f) 90%預測區間的平均寬度，橫軸座標皆為訓練期的天數。

Figure 4. Test of training period lengths for global BMA model. (a) CRPS, (b) MAE, (c) coverage of 66.7% prediction interval, (d) coverage of 90% prediction interval, (e) average width of 66.7% prediction interval, (f) average width of 90% prediction interval, the horizontal axes are the days in training period.

其中以15天為訓練期的結果為最接近，相差僅有0.2%，而以20天為訓練期的結果也僅相差0.3%。圖4(e)和圖4(f)分別是66.7%和90%預測區間的寬度，這兩張圖中發現隨著訓練期越長，預測區間也會

越寬，隱含著若想要得到較精確的預測區間，則不能將訓練期拉得太長。經由上述結果，我們認為針對此模型較合適的訓練期為20天。

在上述的訓練期測試中，我們發現



玉山測站的資料，其觀測值落在90%預測區間外的比例遠大於預期的10%，若忽略玉山測站的資料後重新建立16個測站點上的全區BMA模型，得到的結果如圖5所示。將圖5與圖4比較，可發現忽略玉山資料後建立的BMA模型表現得更好，不僅CRPS跟MAE都明顯地降低，而且66.7%和90%預測區間的涵覆比例都與理想值非常接近，預測區間的寬度也較窄

較精確。因此我們確定玉山的數據並不適合套用在17個測站資料建立出來的全區BMA模型中，把玉山資料除去後BMA模型的預測會較準確。此結果亦讓我們思索是否針對不同地區分別建立不同的BMA模型會比利用全區資料所建的模型更精確？

由於Reftery et al. (2005) 提出的BMA模型建立方法（本文稱之為全區

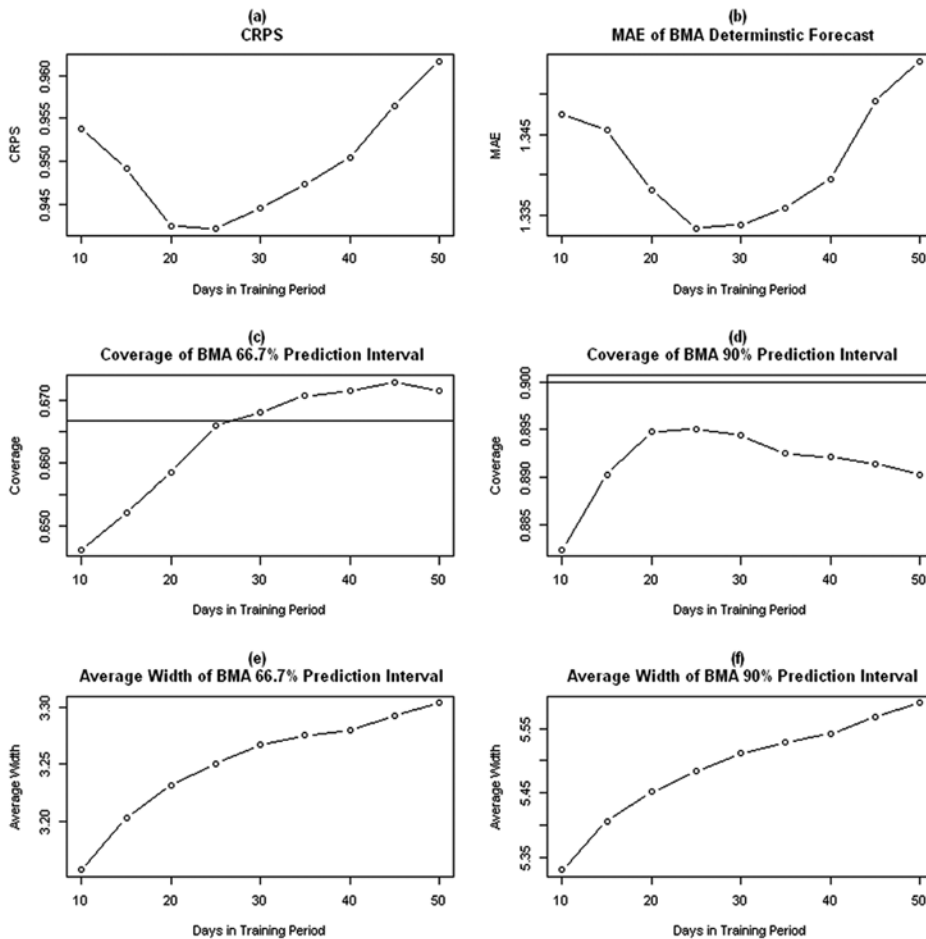


圖5. 同圖4，但忽略玉山測站的資料。

Figure 5. Same as Figure 4, but exclude data from Yushan station.

BMA) 無法隨著不同地區而變，因此我們試著根據不同的測站分別去建立不同的BMA模型(本文稱之為分區BMA)，觀察此方法的表現。為避免模型過於複雜，我們希望不同測站點的BMA模型皆採用相同的訓練期。圖6為測試各種訓練期下分區BMA模型的校驗結果。由於採取分區的方式，每一個測站的BMA模型

只採用該測站的資料來建立，資料筆數相對較少，因此從圖6中可發現分區BMA模型需要較長的訓練期才会有較佳的表現，最佳的訓練期大概介於40天到50天之間，而我們採用45天做為訓練期。圖6中的CRPS和MAE皆比全區BMA低很多，顯示分區BMA模型的確表現更好。但若觀察預測區間的涵覆比例，則發現

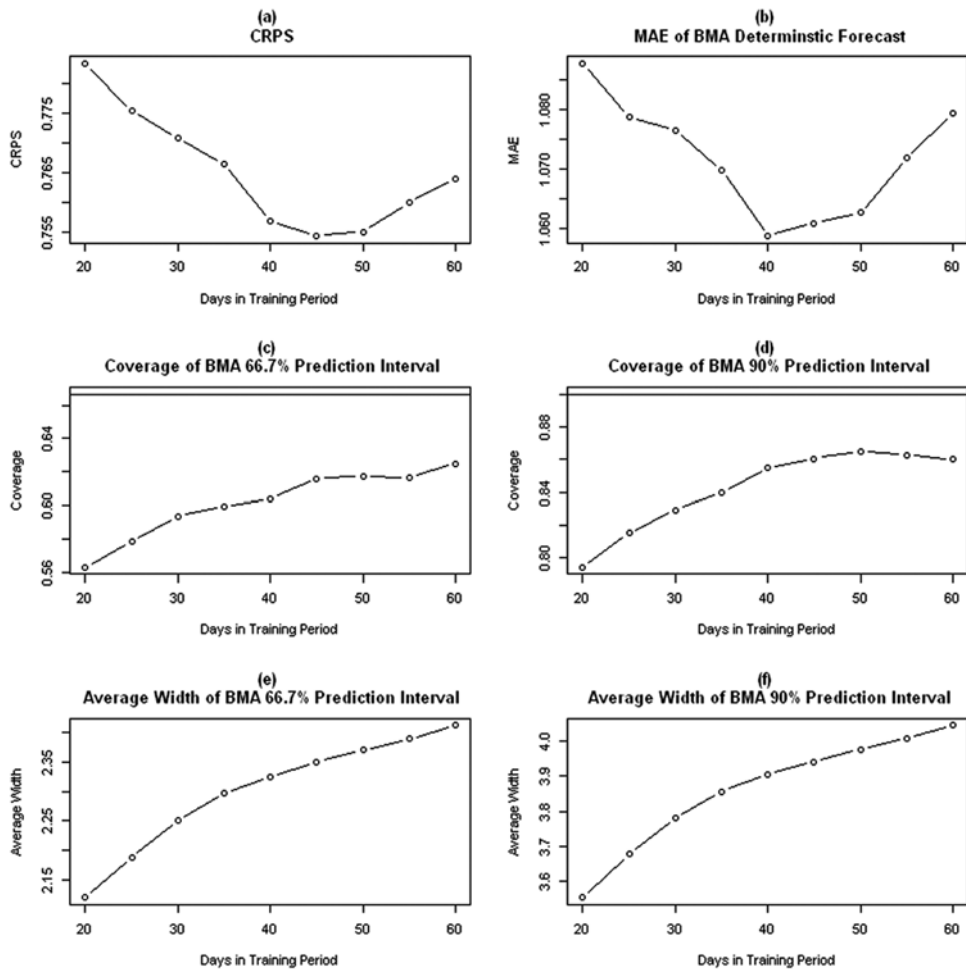


圖6. 分區BMA模型訓練期測試。圖內容解釋與圖4相同。

Figure 6. Test of training period lengths for local BMA model. The contents explain the same as in Figure 4.

比理想值還要低一些，以45天訓練期為例，66.7%預測區間比理想值低了5.1%，而90%預測區間則低了3.9%。雖然分區BMA模型比全區BMA有更狹窄的預測區間，但預測區間的涵覆比例稍顯不足。

### (三) BMA與傳統方法之比較

圖7為RH與PIT校驗圖，此圖可用來比較各種不同方法下的預測機率分布，圖7(a)至(c)代表傳統方法，圖7(d)和(e)則代表BMA方法。其中圖7(a)為使用預報

成員原始預報值的RH圖，圖7(b)和圖7(c)為將預報值進行偏差調整後的RH圖，偏差調整的方式分別為使用全區BMA的調整參數 ( $a_k$ 、 $b_k$ ) 和分區BMA的調整參數 ( $a_{ks}$ 、 $b_{ks}$ ) 進行線性調整，圖7(d)為全區BMA模型的PIT圖，圖7(e)為分區BMA模型的PIT圖，虛線代表直方圖的理想高度。圖7(a)的左側特別突出，代表很多觀測值都低於預報成員的最小值。而經過偏差調整後，顯示觀測值偏低於預報值

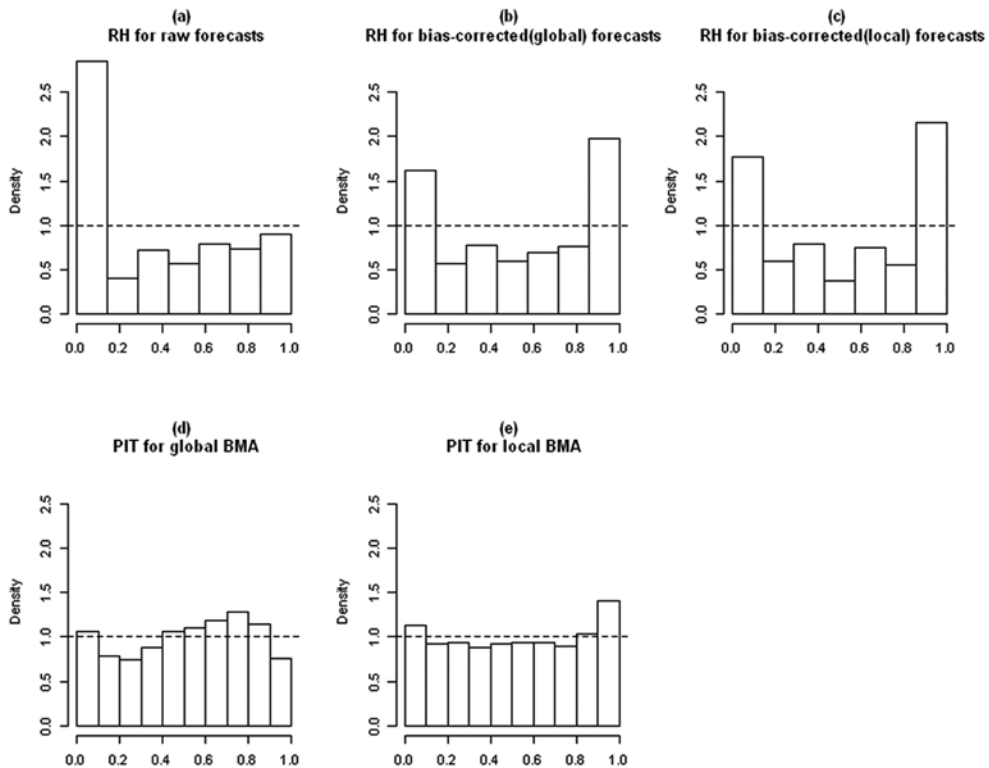


圖7. RH與PIT直方圖。(a)原始預報值的RH圖，(b)全區偏差調整預報值的RH圖，(c)分區偏差調整預報值的RH圖，(d)全區BMA的PIT圖，(e)分區BMA的PIT圖。

Figure 7. RH and PIT histograms. (a) RH for raw forecasts, (b) RH for bias-corrected (global) forecasts, (c) RH for bias-corrected (local) forecasts, (d) PIT for global BMA, (e) PIT for local BMA.

的現象已有改進，但大部分都分散在預報值最大和最小值的區間之外。圖7(a)至(c)反應出，預報成員存在著離散程度不夠的現象（儘管經過偏差調整），無法有效掌握觀測值的分散範圍，因此BPE方法下的預測機率分布與真實狀況不太符合。相較而言，圖7(d)和(e)兩張PIT直方圖都平坦許多，代表BMA的預測機率分布與真實狀況較為接近。其中分區BMA的PIT圖比全區BMA更平坦，唯兩側較高，這反應在圖6中，分區BMA預測區間的涵覆比例低於理想值的現象。

表1為CRPS的校驗結果，在此也比較了五個不同方法的CRPS值，前三個為傳統方法，第一個是預報成員的原始預報值，第二和第三分別為使用全區和分區BMA偏差調整參數調整後的預報值，而第四和第五分別為全區BMA和分區BMA模型的校驗結果。自表中可以發現原始預報值的表現最差而分區BMA模型的表現最好，兩者差距很大。而兩種偏差調整後的預報值其CRPS值也差異頗大，顯示出全區的偏差調整方式較不適合用在臺灣地區，此亦是造成全區BMA模型在本文中表現較差的部分因素。

表2為MAE的比較結果，其結果與CRPS相似。而MAE的比較中，我們另外加入了張等人(2002)發展的系集預報結果進行比較，其採用的方法即是Krishnamurtiet al. (2000)提出的超系集預報（superensemble forecast），然而超系集預報為使用迴歸模型將多組

模式預報合併為決定性預報的多模式迴歸法(multi-model regression method)。儘管加入了超系集預報進行比較，分區BMA模型的期望值依然為最佳的決定性預報。

表1. CRPS的比較結果。

Table 1. Comparison of CRPS.

預報方法	CRPS
Raw forecasts	1.578
Bias-corrected (global) forecasts	1.379
Bias-corrected (local) forecasts	0.879
Global BMA	1.061
Local BMA	0.754

表2. MAE的比較結果。

Table 2. Comparison of MAE.

預報方法	MAE
Raw forecasts	1.900
Bias-corrected (global) forecasts	1.735
Bias-corrected (local) forecasts	1.075
Global BMA	1.475
Local BMA	1.057
Superensemble	1.146

表3為不同方法下預測區間的涵覆比例。對於傳統方法，BPE假設六個預報成員的最小值和最大值代表71.4%預測區間，因此我們取71.4%預測區間的涵覆比例和寬度來對所有方法進行較驗。此外，針對BMA方法也比較了66.7%以及90%預測區間。表中顯示對於傳統方法的預測區間，其涵覆比例遠低於71.4%，

甚至連50%都不到，代表有半數以上的觀測值落在傳統預測區間之外。相較之下，BMA的預測區間與理想值較接近。但如同前面提到的，分區BMA的涵覆比例略低於理想值，且與全區BMA相比，分區BMA與理想值的差距較大。儘管全區BMA的預測PDF其涵覆比例與理想值較接近，圖7中全區BMA的PIT直方圖顯示出其預測PDF與真實狀況的偏差較大。

表4為預測區間的寬度。在涵覆比例差異不大的條件下，預測區間的寬度越窄越好。將表4與表3比較，我們發現三

種傳統方法的預測區間涵覆比例雖然相差不大，但分區偏差調整的預測區間寬度最小。而分區BMA的涵覆比例與全區BMA相比儘管稍顯不足，但其預測區間寬度卻小很多。以71.4%預測區間為例，分區BMA的寬度為2.589而全區BMA為4.120，顯示分區BMA在預報上的表現較精確。此外，傳統方法中分區偏差調整的區間寬度（1.612）雖比分區BMA還小，但由於傳統方法的預報涵覆比例很差（見表3），因此不適合將兩者做比較。

表3. 預測區間的涵覆比例。

Table 3. Coverage of prediction intervals.

預報方法	Coverage		
	66.7%	71.4%	90%
Raw forecasts	NA	46.2%	NA
Bias-corrected (global) forecasts	NA	48.5%	NA
Bias-corrected (local) forecasts	NA	43.8%	NA
Global BMA	69.6%	74.3%	90.3%
Local BMA	61.6%	66.2%	86.1%

表4. 預測區間的平均寬度。

Table 4. Average width of prediction intervals.

預報方法	Width		
	66.7%	71.4%	90%
Raw forecasts	NA	2.949	NA
Bias-corrected (global) forecasts	NA	2.880	NA
Bias-corrected (local) forecasts	NA	1.612	NA
Global BMA	3.736	4.120	6.317
Local BMA	2.350	2.589	3.940

## 五、BMA在氣象預報上的應用

由於透過BMA方法，可以建立出一個完整的預測PDF，因此在氣象預報上的應用有非常大的彈性。除了之前提到BMA可產出預測區間和決定性預報之外，也可設定任意的門檻值，來計算預報因子高於（或低於）此門檻值的極端機率。

圖8為台北測站在2011年5月31日

00UTC的第24小時地面溫度預報，分別使用全區和分區BMA模型所建立的預測PDF。圖中實曲線代表預測PDF，垂直線的部分，細實線為期望值，兩虛線所夾為90%預測區間，粗實線為觀測值，上方圓點則是每一個預報成員的原始預報值。圖中可以觀察出兩種BMA方法下所建立的預測PDF之差異，兩者期望值與觀測值之間的誤差相差均不大，但分區BMA的預測區間寬度卻小很多，顯示分

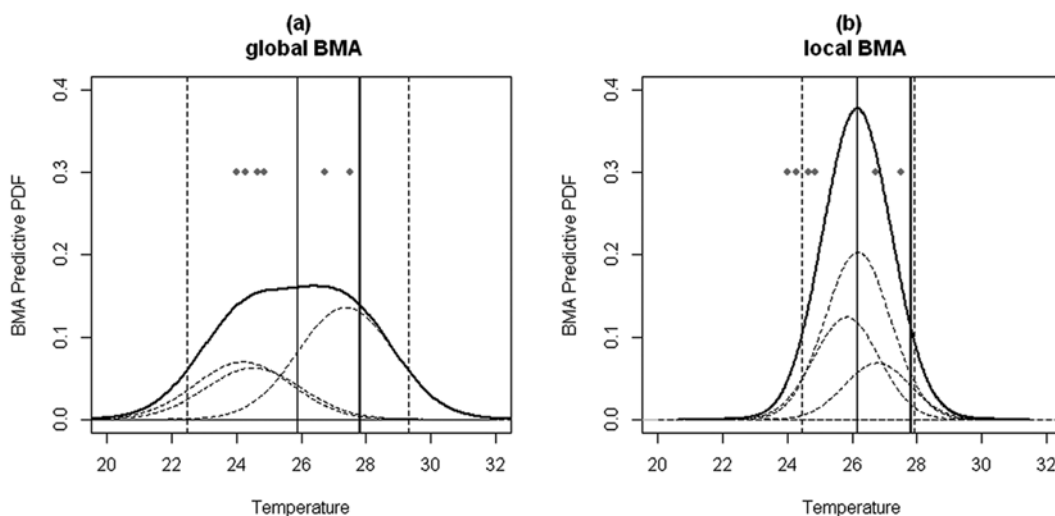


圖8. 2011年5月31日00 UTC之第24小時溫度預報，台北測站的預測PDF。(a) 全區BMA模型下的預測PDF，(b)為分區BMA模型下的預測PDF。實曲線代表BMA方法下的預測PDF，而虛曲線則是每一個預報成員的PDF乘上權重。垂直線的部分，細實線為期望值，虛線為90%預測區間，粗實線為觀測值。上方的圓點則分別代表每一個預報成員的原始預報值。

Figure 8. Predictive PDFs for the 24-h temperature at Taipei station, initialized at 00UTC on 31 May 2011. (a) Global BMA predictive PDF, and (b) local BMA predictive PDF. The thick curve is BMA predictive PDF, and the thin curves are forecast PDFs multiply by weights base on each model. For the vertical lines, the solid thin line is expectation, the dotted lines are 90% prediction interval, and the solid thick line is observation. The upper dots represent the member forecasts.

區BMA在進行預報時的不確定性較低。若依分區BMA對該時間地點進行預報，溫度的期望值為攝氏26.2度，有90%的機率會介於攝氏24.4至27.9度之間，若將攝氏30度視為高溫的門檻值時，預報僅有 $1.4 \times 10^{-4}$ 的機率會發生高溫。

之前曾提及，全區BMA在玉山測站的預報表現特別差，因此我們也針對玉山測站進行BMA的預報結果檢視。圖9為玉山測站的預測PDF，圖中全區BMA與分區BMA的預測區間寬度差異並不大，但經過校驗顯示全區BMA的預測區間並未涵覆住觀測值，而分區BMA的預測區間不僅涵覆觀測值而且期望值與觀測值之間的差異相當小。

由於分區BMA模型的表現較佳，圖10為利用分區BMA將臺灣地區17個測站於2011年5月31日00UTC第24小時溫度預

報的預測PDF。在17個測站中，雖然有8個測站(淡水、臺北、松山、高雄、台中、玉山、新竹、蘭嶼)的原始預報值的範圍皆未涵蓋觀測值，但僅有2個(松山和高雄)測站的90%預測區間未涵蓋。由此可見，儘管在預報成員無法確實涵蓋觀測值的情況下，透過BMA還是能有效掌握觀測值可能的分散範圍。

接著，我們可利用分區BMA模型的預測區間，建立24小時逐時預報的時間序列圖，圖11是利用2011年5月31日00UTC時的24小時逐時預報資料，分別對每小時預報建立一組分區BMA模型(每小時預報的BMA模型訓練期皆設定為45天)。圖中的五條虛線由下到上分別為預測PDF的5%、16.7%、50%(可視為決定性預報)、83.3%、95%分位數，而粗實線為觀測值。透過校驗顯示此預

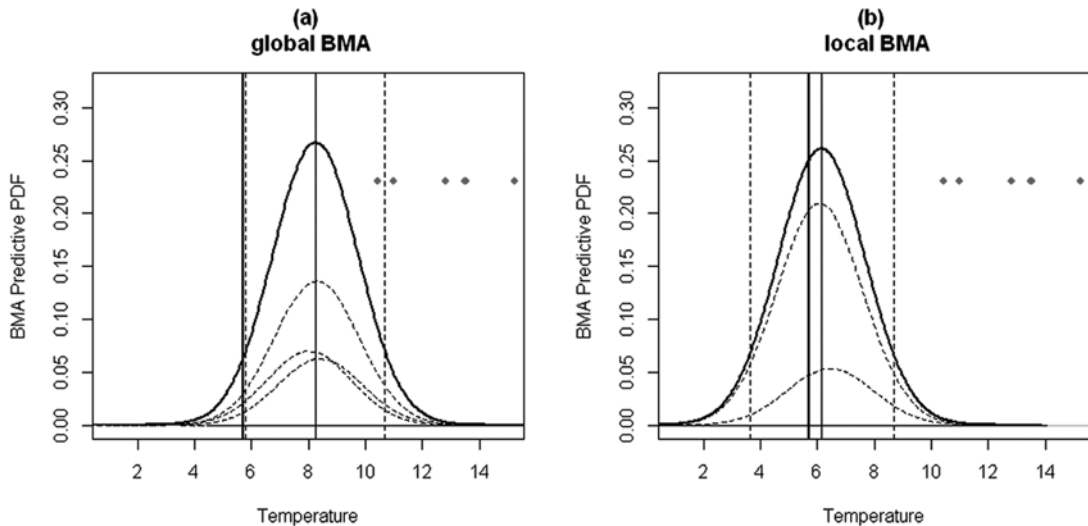


圖9. 同圖8，但為玉山測站。

Figure 9. Same as Figure 8, but at Yushan station.

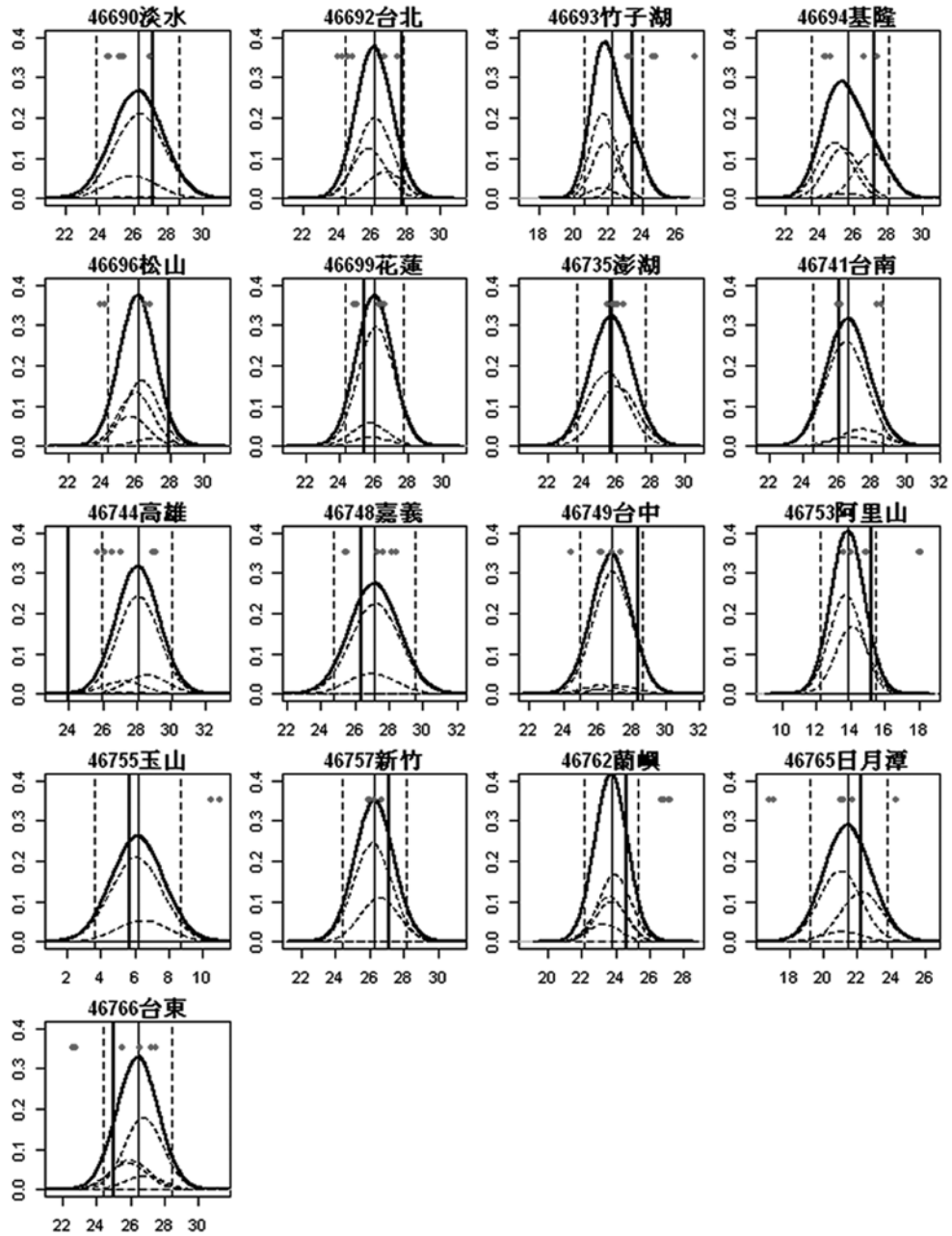


圖10. 2011年5月31日00 UTC之第24小時溫度預報，17個測站在分區BMA模型下的預測PDF。  
圖內容解釋與圖8相同。

Figure 10. Predictive PDFs for the 24-h temperature of 17 stations, initialized at 00UTC on 31 May 2011. The contents explain the same as in Figure 8.



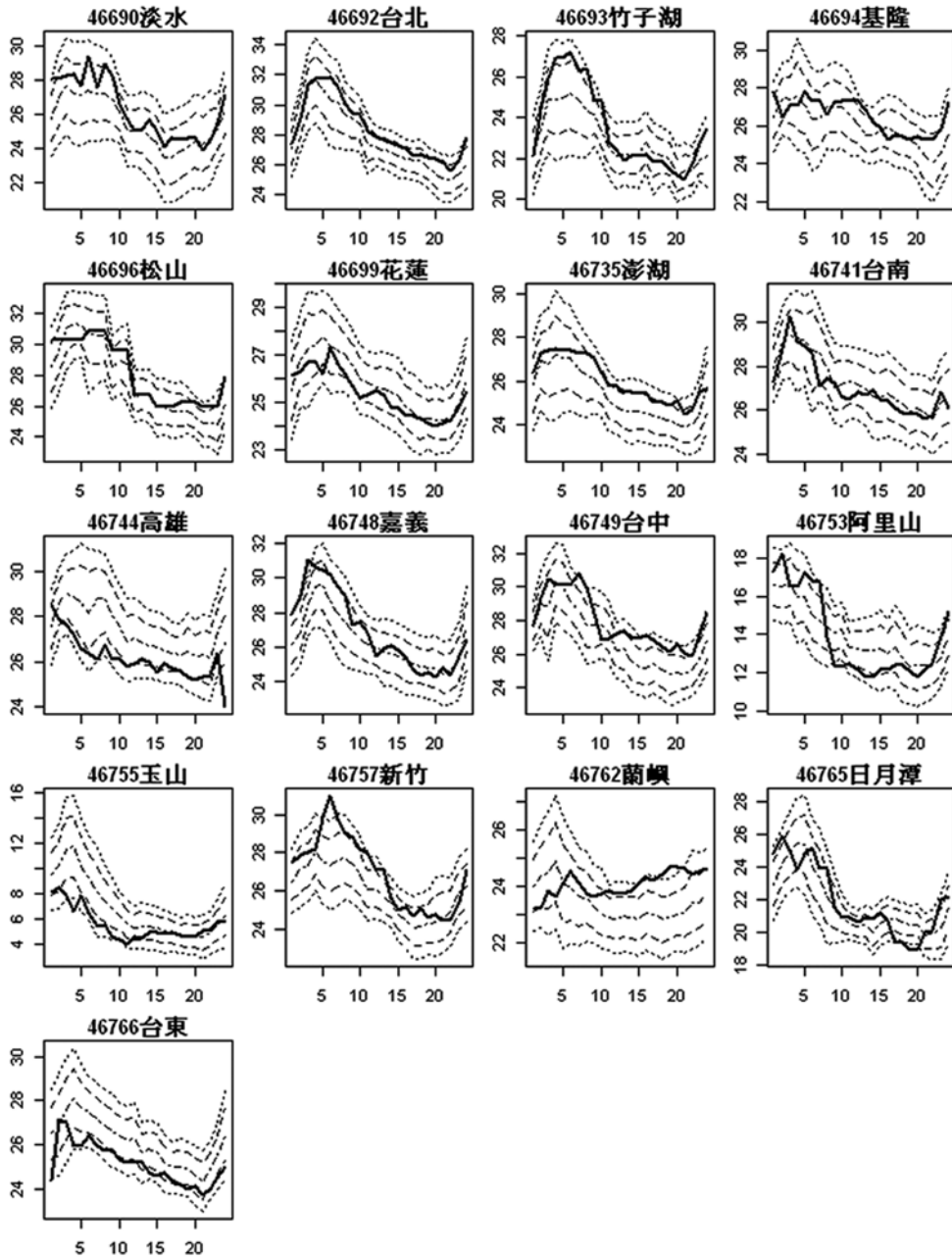


圖11. 2011年5月31日00 UTC之24小時預報時序圖。橫軸為預報小時，圖中的五條虛線分別為預測PDF的5%、16.7%、50%、83.3%、95%分位數連線，粗實線為觀測值。

Figure 11. Time series plots for 24 hours temperature forecast of 17 stations, initialized at 00UTC on 31 May 2011. The horizontal axes are forecast hour, the dotted curves are 5%, 16.7%, 50%, 83.3%, and 95% quantile of the predictive PDFs, and the solid thick curve is observation.

測時序圖對實際溫度變化的掌握能力還算不錯，而藉由此圖可以很容易地預測同一地點未來24小時溫度可能的變化情形。

此外，我們也可設定特定的門檻值，透過BMA模型計算極端天氣的機率，例如夏天設定攝氏35度為熱浪、冬天設定攝氏10度為寒流的警戒值。圖12為採用攝氏35度為高溫門檻值，畫出台北測站於2011年5月31日00UTC時的極端機率時序圖。可將圖12與圖11中的台北站進行對照，兩者為相同的BMA預報但是依不同預報目標而以不同形式來呈現。

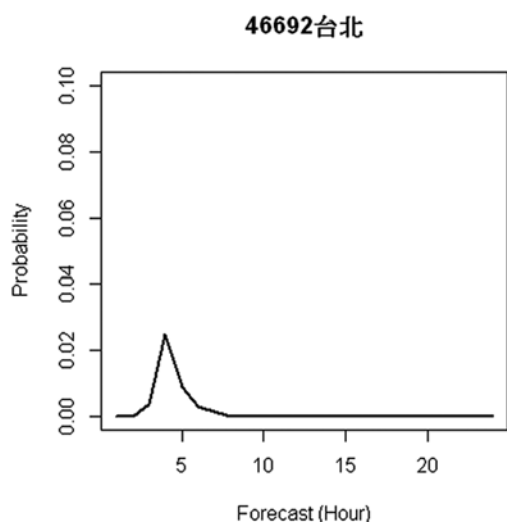


圖12. 2011年5月31日00 UTC於台北測站高溫極端機率之24小時預報時序圖，圖中設定的高溫門檻值為攝氏35度。

Figure 12. Time series plot shows predictive probability for the 24 hours extreme high temperature forecast at Taipei station, initialized at 00UTC on 31 May 2011. The threshold is 35 degree Celsius.

## 六、結論

本文的目的是希望在預報系統中，透過統計方法進行模式預報資料的後製處理，建立起有效的預測機率分布，達到更精確的氣象預報。研究過程採用BMA方法嘗試來達成此目的，並利用臺灣地區測站近地面溫度的觀測資料，進行實際預報的分析與校驗。

BMA的特色除了可以建立起完整的預測PDF之外，BMA給予每一個模式成員不同的權重，有別與傳統方法假設每一個成員的準確度相同。而且BMA亦可以利用假設誤差的機率分布，來解決模式成員離散程度不夠的現象。

本文參考Reftery et al. (2005) 的方法來建立BMA模型，並應用在臺灣地區的溫度資料。再依不同區域分別建立分區BMA模型，與全區BMA進行比較。在經過校驗後，發現分區BMA誤差更低，不確定性也較小。但分區BMA在PIT圖上可以發現它的離散程度稍微不夠，這是未來可以再進行改善的研究方向，或許可以針對 $\sigma$  參數的估計方式做些調整來改善此問題。

整體而言，BMA比起傳統的方法有很大的改進，藉由預測機率分布，BMA能有效的掌握模式預報的不確定性，若能有效運用必能提供更精確的機率預報。

## 七、參考文獻

- 張庭槐，2002：多模式迴歸系集預報系統之建立。交通部中央氣象局專題研究報告第CWB91-1A-10號，22pp。
- Anderson, J. L., 1996: A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *J. Climate*, 9, 1518–1530.
- Fraley, C., A. E. Raftery, T. Gneiting and J. M. Sloughter, ensembleBMA: An R Package for Probabilistic Forecasting using Ensembles and Bayesian Model Averaging, Technical Report No. 516R, Department of Statistics, University of Washington, 2007 (revised 2010).
- Gneiting, T., A. E. Raftery, A. H. Westveld, and T. Goldman, 2005: Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Mon. Wea. Rev.*, 133, 1098–1118.
- Hersbach, H., 2000: Decomposition of the continuous ranked probability score for ensemble prediction systems. *Wea. Forecasting*, 15, 559–570.
- Hoeting, J. A., D. M. Madigan, A. E. Raftery, and C. T. Volinsky, 1999: Bayesian model averaging: A tutorial (with discussion). *Stat. Sci.*, 14, 382–401. [A corrected version is available online at [www.stat.washington.edu/www/research/online/hoeting1999.pdf](http://www.stat.washington.edu/www/research/online/hoeting1999.pdf).]
- Krishnamurti, T. N., Z Zhang, T. E. LaRow, D. R. Bachiochi, C. E. Willifor, S. Gadgil, and S. Surendran, 2000: Multi-model ensemble forecasts for weather and seasonal climate, *J. Climate*, 13, 4196–4216.
- 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.
- Sloughter, J. M., A. E. Raftery, T. Gneiting, and C. Fraley, 2007: Probabilistic quantitative precipitation forecasting using Bayesian model averaging. *Mon. Wea. Rev.*, 135, 3209–3220.
- , T. Gneiting, and A. E. Raftery, 2010: Probabilistic wind speed forecasting using ensembles and Bayesian model averaging. *J. Amer. Stat. Assoc.*, 105, 25–35.

# Applying Bayesian Model Averaging to Temperature Probability Forecasts over Taiwan

Yu-Hsuan Chang, Ting-Huai Chang, Huey-Ru Wu  
Meteorological Information Center,  
Central Weather Bureau

## ABSTRACT

Ensemble forecasts aim at addressing the uncertainty of weather scenarios to the best extent possible so as to seek the probable diverse range of future weather. If probability distributions of actual weather are available for reference, this uncertainty can be more accurately described.

First, this research applies Bayesian model averaging (BMA), a statistical method, to the post-processing on multi-sets of model predictions in order to set up a probability density function (PDF) conforming to real weather conditions.

Subsequently, BMA is employed on surface temperature forecasts of the weather stations around Taiwan area. Verification results indicate that PDFs by BMA perform ideally on the forecasts, and even though the dispersion of the predictions of multi-model forecast members is limited, BMA can still effectively capture the uncertainty of temperature variations to a certain degree.

**Key words:** Bayesian model averaging, probability density function, ensemble forecast