

應用貝氏模型平均法發展颱風路徑機率預報指引

陳昱璁¹ 馮智勇¹ 張博雄² 許乃寧² 賈愛玫²

¹多采科技有限公司

²交通部中央氣象局

(中華民國 106 年 10 月 17 日收稿，中華民國 107 年 8 月 29 日定稿)

摘 要

隨著數值天氣預報模式發展與電腦運算能力提升，颱風路徑預報的平均誤差不斷降低，預報不確定性的資訊也能藉由多模式決定性預報或系集模式成員預報的分布情形加以估計，使得發展颱風路徑機率預報更為可行。本研究嘗試將颱風路徑分為經度預報與緯度預報，再分別應用貝氏模型平均法(Bayesian Model Averaging)整合不同模式預報資訊得到經度與緯度上的預報機率密度函數，除可依此建構完整的颱風路徑機率預報指引外，亦可利用期望值作為路徑預報指引。

案例分析使用 2016 年美國 NCEP (National Centers for Environmental Prediction)、歐洲 ECMWF (European Centre for Medium Range Weather Forecasts)、日本 JMA (Japan Meteorological Agency)與中央氣象局 TWRF (Typhoon WRF)模式的颱風路徑預報資料與官方最佳路徑資料，並以等權重平均法作為對照組進行比較。結果顯示貝氏模型平均法的颱風路徑期望值預報在 12、24、36 與 48 小時預報誤差顯著優於等權重平均，24 與 48 小時的預報誤差可分別降低至約 65 公里與 110 公里。進一步區分為緯度預報與經度預報誤差分析可知，由於四個模式颱風中心位置的緯度預報差異性較大，有利於貝氏模型平均法依據模式預報能力給予不同權重係數而能優於等權重平均者，使得期望值預報誤差改善幅度明顯較經度預報結果為高。另一方面，颱風路徑機率預報同樣因貝氏模型平均法具有區別模式預報能力的特性，可針對四個模式颱風中心位置預報分散的情況提供各種路徑發生機率的高低差異而能輔助預報作業人員修正期望值預報指引。

關鍵字：颱風路徑機率，貝氏模型平均法

一、前言

颱風路徑預報技術近年來受惠於資料同化、動力數值預報模式能力改善、系集模式興起與統計後處理方法改進而不斷提升 (Elsberry, 2014)，中央氣象局 24 小時颱風中心預報誤差平均值自 2010 年起已低於 100 公里，並且於 2016 年減少到約 77 公里。隨著各國氣象作業單位與研究中心致力發展系集模式以及無縫隙預報概念興起，預報作業人員除已難以於有限時間評估各種模式結果的預報能力外，亦僅能主觀利用自身經驗判斷而不易維持預報品質。因此，如何客觀整合多模式或者系集模式預報結果為進一步降低颱風路徑預報誤差的重要課題。

Elsberry et al. (2008) 提出 Weighted Position Consensus 與 Weighted Motion Vector Consensus 兩種將颱風路徑拆解為經度預報與緯度預報而整合多模式預報結果的方式，此方法係假設預報作業時間與模式模擬起始時間(initial time)距離 72 小時，則可透過觀測路徑誤差的距離平方反比計算各個模式的權重，並且將這個權重套用到 72 小時後的預報。Weighted Position Consensus 是對預報位置做加權平均，Weighted Motion Vector Consensus 則是將模式從 72 小時後每單位時間的移動向量做加權平均，並且從颱風中心觀測位置延伸整合預報。Jun et al. (2015) 利用 Superensemble 法 (Krishnamurti et al., 2000; Williford et al., 2003) 整合多種系集模式於西

北太平洋區域的熱帶氣旋路徑，在 24、48、72、96、120 小時的預報中較等權重平均法分別獲得 12.4%、10.1%、9.9%、10.8%、16.5% 的改善幅度。Dong et al. (2016) 同樣基於預報作業時間與模式模擬起始時間具有一定時間延遲的特性，提出從系集模式中抽出較接近觀測的子集合的 Observation- Based Ensemble Subsetting Technique (OBEST) 方法，使得 12~36 小時內的颱風路徑預報誤差顯著優於系集成員平均值。

考量颱風路徑預報不確定性隨著預報時間增加，預報作業人員需透過各種預報資訊掌握颱風路徑可能的變異範圍以示警可能受影響區域。中央氣象局為了量化路徑預報不確定性，自 2007 年起提供颱風路徑侵襲機率產品 (Probabilities of Strike, POS; 顧 2006、蔡 2007)，蔡等人(2011)進一步利用官方預報路徑為依據區分為徑向(along track)與橫向(cross track)，並且以蒙地卡羅法(Monte Carlo Method)亂數隨機挑選歷史颱風預報誤差的方式發展西北太平洋颱風侵襲臺灣機率的估計方式。此方法必須先行給定颱風路徑預報後再利用長期統計資料擬合誤差的常態分佈參數，而非直接整合自各種模式最新預報資料，若是近期有模式改版而造成誤差型態有所變化，則可能會有錯估的情況發生。

上述方法可單獨獲得颱風路徑預報，並藉由長期統計資料額外給予預報誤差估計資訊，或者是藉由系集成員估計各種發生機率，

若能透過分析近期各種模式預報能力並轉為機率形式而予以整合，則可於獲得機率預報的同時也得到一組決定性預報，除有利於直接比較各種方法產製的路徑預報優劣外，更可提供作業人員掌握路徑可能變化趨勢而進行主觀修正決策或後續防災參考。

本研究嘗試應用貝氏模型平均法 (Bayesian Model Averaging, BMA) (Leamer 1978; Kass and Raftery, 1995; Hoeting et al., 1999) 發展客觀整合多模式颱風路徑預報的方式。Roberts (1965) 提出由兩個模型以後驗機率 (posterior probability) 加權平均得到新模型，為 BMA 法的雛型，並由 Leamer (1978) 予以延伸而提出基本架構。Hoeting et al. (1999) 提出將所有可能的模型以後驗機率做加權平均後得到混合模型，再以該混合後的模型來進行預測的方法。然而，計算機運算能力以及較為複雜的後驗機率估算方式，始終為 BMA 應用上的障礙。直到 Raftery et al. (2005) 提出利用事先分析條件機率密度函數形式以簡化計算得實作方式，並使用華盛頓大學所模擬的 Fifth-Generation Pennsylvania State University-NCAR Mesoscale Model (MM5) 系集模式改善美國 2000 年 1 月至 6 月的美國西北地區 (Pacific Northwest) 地面溫度預報後，BMA 方廣為氣象與氣候領域所應用，亦有實作於臺灣測站溫度機率預報指引的應用 (張等人, 2011; 馮等人, 2015)。

本研究參考 Raftery et al. (2005) 實作方

式，發展客觀整合多模式颱風路徑預報資訊而得中心位置機率密度函數的作業流程，並統計校驗以期望值與機率分布的正確性。考量颱風行徑路線為二維空間事件，因此參考 Elsberry et al. (2008) 作法將其拆解為經度預報與緯度預報再進行。以下依序說明 BMA 法架構、使用資料、模式路徑預報條件機率分布、參數敏感度分析與校驗分析結果，最後以個案呈現颱風路徑機率預報使用結果與研究結論。

二、貝氏模型平均法颱風路徑機率預報指引建置

(一) 貝氏模型平均法架構

BMA 算式如下式(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 y 基於已知成員 k 的預報值 f_k 時的條件機率分布， $p(f_k | y^T)$ 則是在訓練資料 y^T 之中 f_k 的表現為成員中最佳者的後驗機率，兩者相乘後加總，就能得到等號左側由各模式整合後的物理量 y 的發生機率分布預測。

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

Raftery et al. (2005) 假設可將條件機率分布以特定機率密度函數代表，並且後驗機率滿足 $\sum_{k=1}^K p(f_k | y^T) = 1$ 可視為權重係數 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一組權重係數，描述該機率密度函數所需的參數的數量與種類要視誤差分布而定。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 \mathcal{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標準差 σ_k ，可藉由最大似然函數(Maximum likelihood)法(Fisher 1922)進行估計。

$$p(y|f_1, f_2, \dots, f_K) = \sum_{k=1}^K w_k g(y|f_k, \sigma_k) \quad (2)$$

似然函數(likelihood function)為訓練期資料代入不同參數時能夠命中觀測數值的機率，而最大似然函數法的目標就是估計出一組最佳參數，使得訓練期的預報命中觀測資料機率為最大。Raftery et al. (2005)為了計算方便與數值穩定，將似然函數改為對數似然函數(log-likelihood function)如式(3)所示，並以 Expectation Maximization algorithm (EM 演算法) (Dempster et al., 1977; McLachlan and Krishnan, 1997)求解。其中， t 代表不同時間的觀測與預報配對，每一筆訓練資料 t 包含了時間相對應的觀測與所有預報成員。

$$l(w_1, w_2, \dots, w_K, \sigma_1^2, \sigma_2^2, \dots, \sigma_K^2) = \sum_t \log \left(\sum_{k=1}^K w_k g(y_t | f_{kt}, \sigma_k) \right) \quad (3)$$

EM 演算法分為兩個步驟，第一個步驟

(E 步驟)先設定一個潛在變數 z_{kt} 計算似然函數估計值，也就是藉由該潛在變數作為當前參數帶入似然函數時各個成員 k 在各組訓練配對 t 的表現依據，並且在第二個步驟(M 步驟)藉由該潛在變數的資訊將變數朝向能最大化似然函數估計值的方向調整待估參數，接著再進入下一輪的 E 步驟。對數似然函數值於迭代過程中會不斷上升(Dempster et al., 1977)，直到數值於兩次迭代變化趨緩而收斂時，即可獲得最佳參數解。

在 E 步驟中引進潛在變數的用意在於分析每組配對資料中的成員表現，標示最佳預報成員為 1，其餘成員則等於 0，而其估計值 \hat{z}_{kt} 計算式如式(4)所示，分子計算每組配對資料中各個成員發生觀測值的機率值 $g(y_t | f_{kt}, \sigma_k^{(j-1)})$ 與權重 $w_k^{(j-1)}$ 的乘積，做為該成員在該組配對資料的得分，並且分母以各成員得分總和進行標準化而介於 0 到 1 間，並用以代入式(5)與(6)計算下一次迭代中的 w_k 與 σ_k 數值，其中， n 代表訓練資料 t 的總數。式(5)中， w_k 來自於 \hat{z}_{kt} 中該成員所有訓練資料預報成效的平均值，而式(6)之中每個成員的標準差平方則是來自於每筆誤差平方乘上 \hat{z}_{kt} 做加權加總後再除以 $\sum_i \hat{z}_{kt}^{(j)}$ 為標準化。觀察式(4)~式(6)可知，預報成員的最佳預報得分平均 \hat{z}_{kt} 會成為影響權重 w_k ，使得表現較佳者的權重能持續放大；式(6)的

誤差標準差計算則是於分子乘上 $\hat{z}_{kt}^{(j)}$ ，試圖藉由放大常態分布標準差而降低觀測值的發生機率 ($g(y_t | f_{kt}, \sigma_k^{(j-1)})$)，抑制表現較佳成員的權重放大，因此可讓各預報成員的權重 w_k 達成平衡。過程中，對數似然函數會不斷增加並變化趨緩，在變化趨近於零時收斂並且完成參數的估計。

$$\hat{z}_{kt}^{(j)} = \frac{w_k^{(j-1)} g(y_t | f_{kt}, \sigma_k^{(j-1)})}{\sum_{i=1}^K w_i^{(j-1)} g(y_t | f_{it}, \sigma_k^{(j-1)})} \quad (4)$$

$$w_k^{(j)} = \frac{1}{n} \sum_t \hat{z}_{kt}^{(j)} \quad (5)$$

$$\sigma_k^{2(j)} = \frac{\sum_t \hat{z}_{kt}^{(j)} (y_t - f_{kt})^2}{\sum_t \hat{z}_{kt}^{(j)}} \quad (6)$$

啟動 EM 演算法(j=1)需要給定參數初始猜值，本研究分別設定 w_k 與 σ_k 的初始值為 $1/K$

以及各預報成員歷史資料的預報誤差標準差。訓練完的各成員權重與標準差，配合新的模式預報，將數值帶入式子(2)之後可以得到一個在空間上連續的機率預報，如圖 1 所示，圖中所呈現的是 BMA 預報機率在緯度上的分布情形，橫軸為緯度，縱軸為機率密度，時間為 2016/07/04-12:00 UTC 預報 048 小時候的結果，預報對象為尼伯特颱風，黑色的線段為 BMA 整合後的機率密度隨緯度變化情形，也就是 $\sum_{k=1}^K w_k g(y | f_k, \sigma_k)$ ，其餘四個線段為各成員的機率預報 $w_k g(y | f_k, \sigma_k)$ ，紅色的星號為 BMA 機率預報的期望值位置，藍綠色的星號為觀測結果。

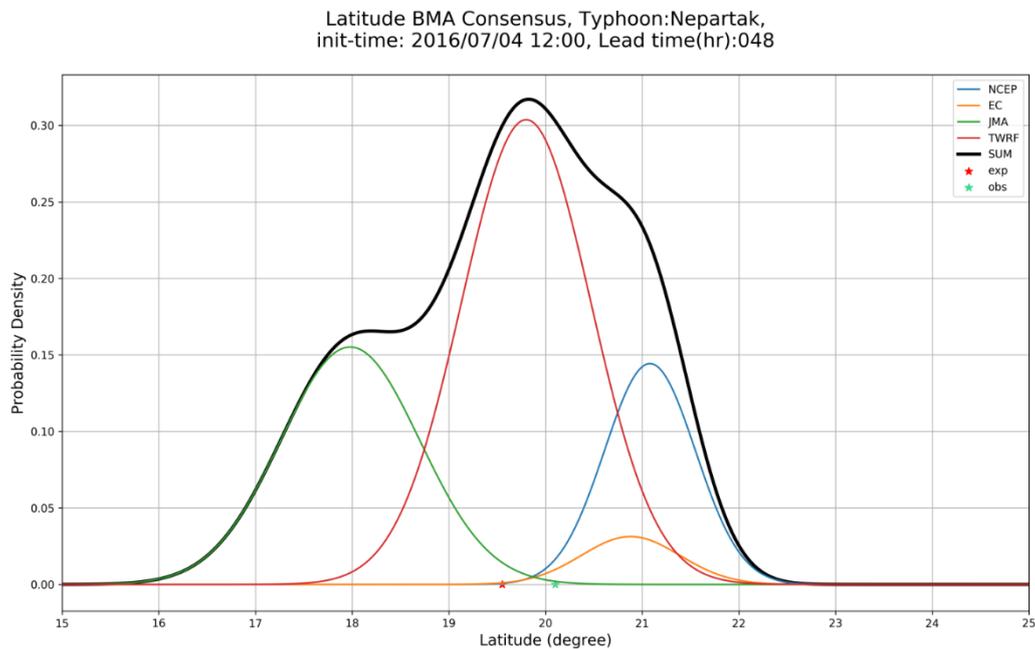


圖 1. BMA 機率預報分布與各成員分布圖，橫軸是緯度，縱軸是機率密度，圖中的黑色線段是整合後的 BMA 預報機率密度隨緯度變化，模式起始時間為 2016/07/04 12:00 UTC，預報目標是 048 小時候的尼伯特颱風路徑，其他線段則是不同成員的預報機率分布，底下的紅色星號代表 BMA 預報的期望值位置，藍綠色的星號代表觀測結果。

由於條件機率是基於模式預報結果分析與觀測值誤差特性，因此當條件機率符合常態分布機率密度函數特性時，標準差可視為模式預報能力，標準差越大則能力越差，應給予較小。EM 演算法迭代過程中，常態分布的標準差是由式(6)計算而得，允許微調以利得到最大似然函數值。由式中可看出每一個誤差平方後會乘上對應的 \hat{z}_{kt} ，並且加總後除以該模式的 \hat{z}_{kt} 加總進行標準化，與直接計算誤差標準差意義相近。考慮到 \hat{z}_{kt} 的性質，在該次 t 預報時表現好的(誤差相對其他成員小的)可以得到較高的 \hat{z}_{kt} ，也就是說預報表現相對其他成員而言較好的幾次會得到較大的加權，因此嘗試將標準差略為放大，朝向增加次回迭代式(4)中的條件機率值，所以完成迭代後的標準差會接近原始標準差，但是相對大小不確定。

(二) 資料說明

本研究使用的颱風路徑觀測為中央氣象局的颱風最佳路徑(Best Track)資料，預報成員資料則為中央氣象局現行颱風路徑預報作業所參考的美國國家環境預報中心(National Centers for Environmental Prediction，以下簡

稱 NCEP)、歐盟中期天氣預報中心(European Centre for Medium-Range Weather Forecasts，以下簡稱 EC)與日本氣象廳(Japan Meteorological Agency，以下簡稱 JMA)分別發展的全球氣象數值模式以考量颱風移動路徑受到太平洋副熱帶高氣壓的影響，亦納入中央氣象局改進了颱風渦旋初始化流程(Hsiao et al. 2010)、三維變分同化系統背景誤差、Kain-Fritsch 積雨雲參數化(Hsiao et al. 2011)的區域氣象模式 Typhoon WRF(以下簡稱 TWRF)，以掌握颱風受到臺灣地形干擾的效應。

四個模式解析度分別為 NCEP $0.5^\circ \times 0.5^\circ$ 、EC $0.5^\circ \times 0.5^\circ$ 、JMA $0.5^\circ \times 0.5^\circ$ 與 TWRF $3 \text{ km} \times 3 \text{ km}$ ，而本研究中所使用之各模式預報路徑資料是由中央氣象局進行颱風中心定位而得。

EC 模式在 00:00 UTC 與 12:00 UTC 時各會啟動 1 次，其餘則是每日 4 次，因此可用資料為颱風發生時 00:00 UTC 與 12:00 UTC 的模式預報。2015 年與 2016 年的四個預報成員同時具有預報資料的空間分布如圖 2 所示，圖上的點代表模式啟動時的颱風位置，並以 NCEP 預報 000 小時的位置為代表，經度範圍大約是東經 105~175 度，緯度範圍則大約是 0 度至北緯 50 度。

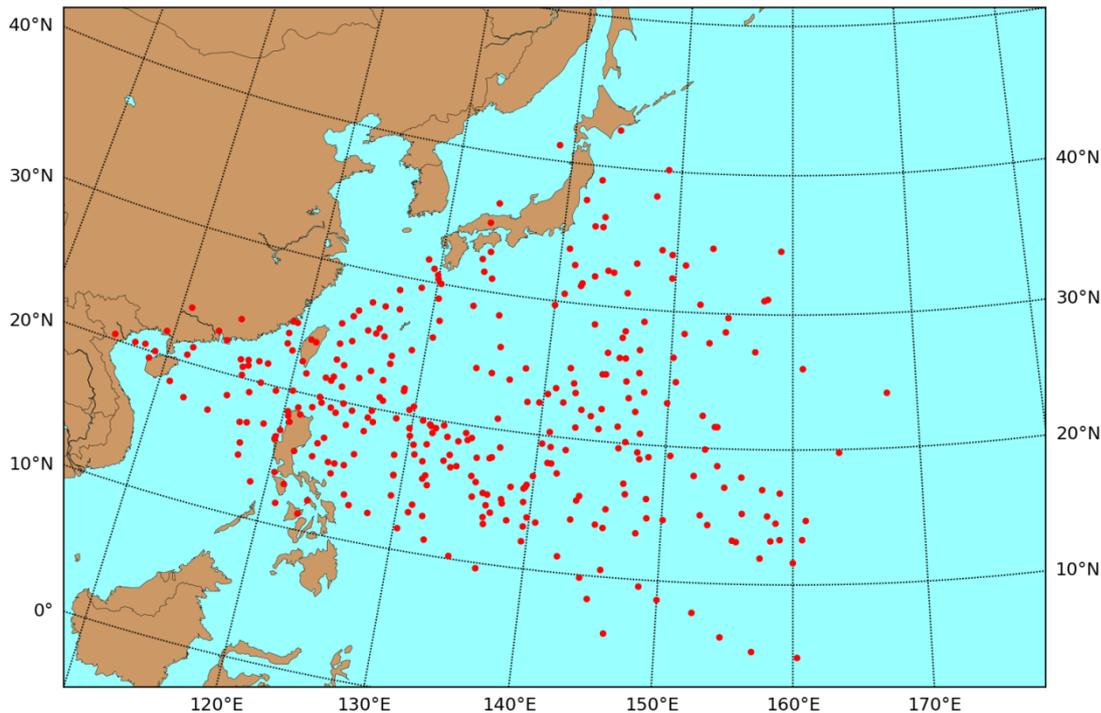


圖 2. 2015 到 2016 年期間 NCEP、ECMWF、JMA 與中央氣象局 TWRf 模式同時存在時，以 NCEP 模式在預報時間 000 所在位置進行標示。

(三) 模式路徑預報的條件機率分析

使用 2015 到 2016 年間數值模式颱風路徑預報於經度預報與緯度預報的預報誤差如圖 3 與圖 4 所示，由左至右各欄依序代表使用 NCEP、EC、JMA 與 TWRf 四種模式，而由上而下依序為預報 012、024、036、048、060、072、084 與 096 小時路徑的直方圖。直方圖是先將誤差資料以路徑預報誤差平均值搭配上 2.5 倍的標準差簡易篩選離群值後，再以橫軸為誤差(單位為度)，縱軸為發生次數繪製而成。為了避免受到 2015~2016 年間零星發生的不良颱風案例資料影響，於此反覆利用選取資料的平均值與 2.5 倍標準差所構成的區間剔除離群值，直到沒有任何資料

樣本被剔除為止。此處的不良颱風案例，主要來自於對模式進行颱風中心定位時的錯誤，可能是上游資料處理或其他問題所造成，與模式本身無關，因此必須要設法找出並剔除。此外，TWRf 因少有預報 096 小時的案例資料而不納入分析。圖中可見四個模式於不同預報時間呈現類似常態分布的特性，因此假設誤差分布符合常態分布而將平均值與標準差進行 Kolmogorov-Smirnov 檢定，並且同時將 P-Value 值(右上方)大於 0.05 的直方圖標上綠色，代表誤差分布符合常態分布；而小於 0.05 且大於 0.01 的則是標上黃色，代表誤差資料非為常態分布，各圖中間的數字代表該組誤差的標準差。

在表示經度預報誤差分布的圖 3 中，可

以看到多數的誤差分布均符合常態分布，僅 EC 模式在預報 060 小時的 P-Value 為 0.025 而被判定為非常態分布，從圖中可以看到此組誤差分布相較於理想的常態分佈而言，其數值更加集中在中央正負 1 度的範圍內，而在此範圍之外的區域則是突然驟降，可能為無法通過檢定的原因。在表示緯度預報誤差分布的圖 4 中，僅 TWRF 模式於預報 012 小時以及 EC 模式於預報 036 小時不符合常態分布的假設。其中，TWRF 的預報 012 小時直方圖可能因兩個區間的分布數量為 0 無法通過統計檢定，而 EC 模式的預報 036 小時則可能因在 -1 度到 -3 度區間的樣本數量少於 1 度至 3 度的樣本數量，並且資料多數集中在中間 -1 度到 1 度的範圍內而無法通過檢定。整體而言，由於路徑誤差於經度預報和緯度預報的分布多數符合常態分布，因此本研究假設已知預報值時的觀測值條件機率 ($g(y|f_k, \sigma_k)$) 為常態分布，逐預報時間分別建置經度預報以及緯度預報的 BMA 模型。我們認為 EC 的預報在這段時期有幾個預報時間沒通過常態檢定，是來自於 EC 模式在 2015 年末到 2016 年中的數次改動造成，而 TWRF 未過檢定的部分推測是與地形處理較精緻導致海上颱風與觸陸颱風的行為稍有差異所導致。

預報時期為了避免受到不同時期的模式預報能力差異所影響，應盡可能使用接近整合當下模式預報資料進行 BMA 模型參數估

計。假設要整合的預報是某年的 8 月 1 號 00:00 UTC，並且想利用 40 筆觀測與模式預報的配對作為訓練期資料，最理想的情況是在該年的 7 月就能湊齊足夠筆數，若無法則只能使用如六月、五月等早期資料。以建置 T 日預報為例，本研究若欲使用 N 筆訓練期資料估計 BMA 模型參數，將從預報時間向前搜尋 5N 筆資料並且計算各模式成員的預報誤差標準差，並依此剔除誤差在正負 3 倍標準差以外的資料，再從剩餘的資料中選出時間最接近 T 日預報時間的 N 筆作為訓練期資料。考量預報誤差行為可能隨時間有些許差異，為了確保近期資料不會因較嚴格的條件而被剔除，因此選擇 3 倍標準差為門檻。本研究採用 NCEP、EC、JMA 與 TWRF 四個模式建置預報 012~084 小時的 BMA 模型，而建置預報 096、108 與 120 小時者則採用 NCEP、EC 與 JMA 三個模式資料，在接下來的各項測試中皆以 2016 年全部颱風案例進行 BMA 模型校驗分析。

(四) 訓練期資料長度敏感度測試

於此以預報的平均誤差做為比較依據，測試訓練期資料長度對於 BMA 模型預報平均誤差的影響。選擇訓練期資料時為了避免空間相距過遠的颱風路徑誤差有著不同的性質，僅將與該次預報的經緯度預報差距在 20 度以內的模式資料納入。結果如圖 5 所示，橫軸代表不同的預報時，而縱軸則是代表各組預報的路徑誤差平均值，不同顏色的直方

圖代表用 30、40、50、60、70 筆資料訓練資料建置的 BMA 模型。從圖中可以看到在 012 與 024 小時預報幾乎不受資料筆數影響，036 小時預報以 50 與 70 訓練筆數的結果略佳，在 048 小時則以 60 筆與 70 筆誤差較小，而 060 小時預報則是 60 筆訓練資料的誤差為最小，並且在預報 036~060 小時的誤差平均值並沒有特定的變化趨勢，主要的共通點為 30 筆訓練資料所得到的結果相對誤差較大。而 072 小時預報的最低誤差訓練筆數為 50 筆，兩側的 30 筆與 70 筆是誤差相對較高的組別，084 小時預報則有類似現象並且以 40~60 筆訓練資料所得結果較為穩定。從圖中可歸納訓練期資料筆數約在 40~50 筆時可以得到較佳結果，因此後續分析中固定訓練期資料筆

數為 40 筆。

(五) 訓練期資料空間分布限制敏感度測試

考慮模式在不同空間位置的預報表現可能有所差異，因此設計空間分布限制參數篩選訓練期資料，此部分的訓練期資料固定為 40 筆資料。空間分布限制的測試設定包括 15 度、20 度、25 度、30 度，設定的度數代表該次預報整合只納入歷史資料的預報路徑回推到 000 小時的預報落在本次 000 小時預報位置的經緯度預報正負 N 度之內，此要求越嚴苛，便越有機會因為待選訓練資料超出空間限制範圍而剔除，造成使用更早期的資料進行訓練，而若限制放寬則會納入許多空間距離較遠的資料做為 BMA 訓練資料。

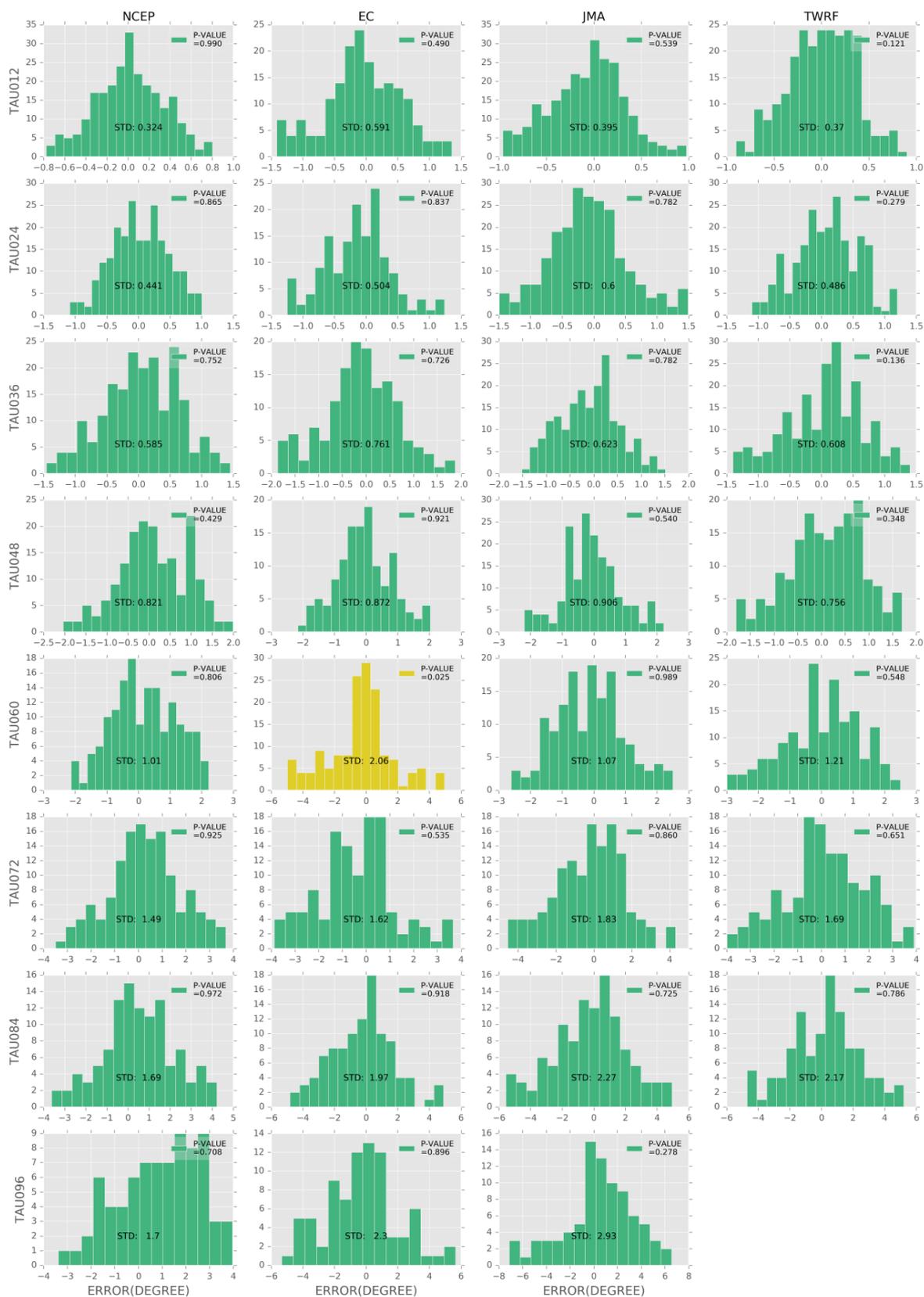


圖 3. 模式在不同預報時間(TAU)的經度預報誤差分布情形，橫軸單位為度數，縱軸單位為發生次數。在右上角為 Kolmogorov-Smirnov 檢驗該分布是否為常態的 P-Value。柱狀圖的顏色為綠色代表 P-Value 大於 0.05，小於 0.05 以黃色表示，各圖中央的數字代表標準

差。

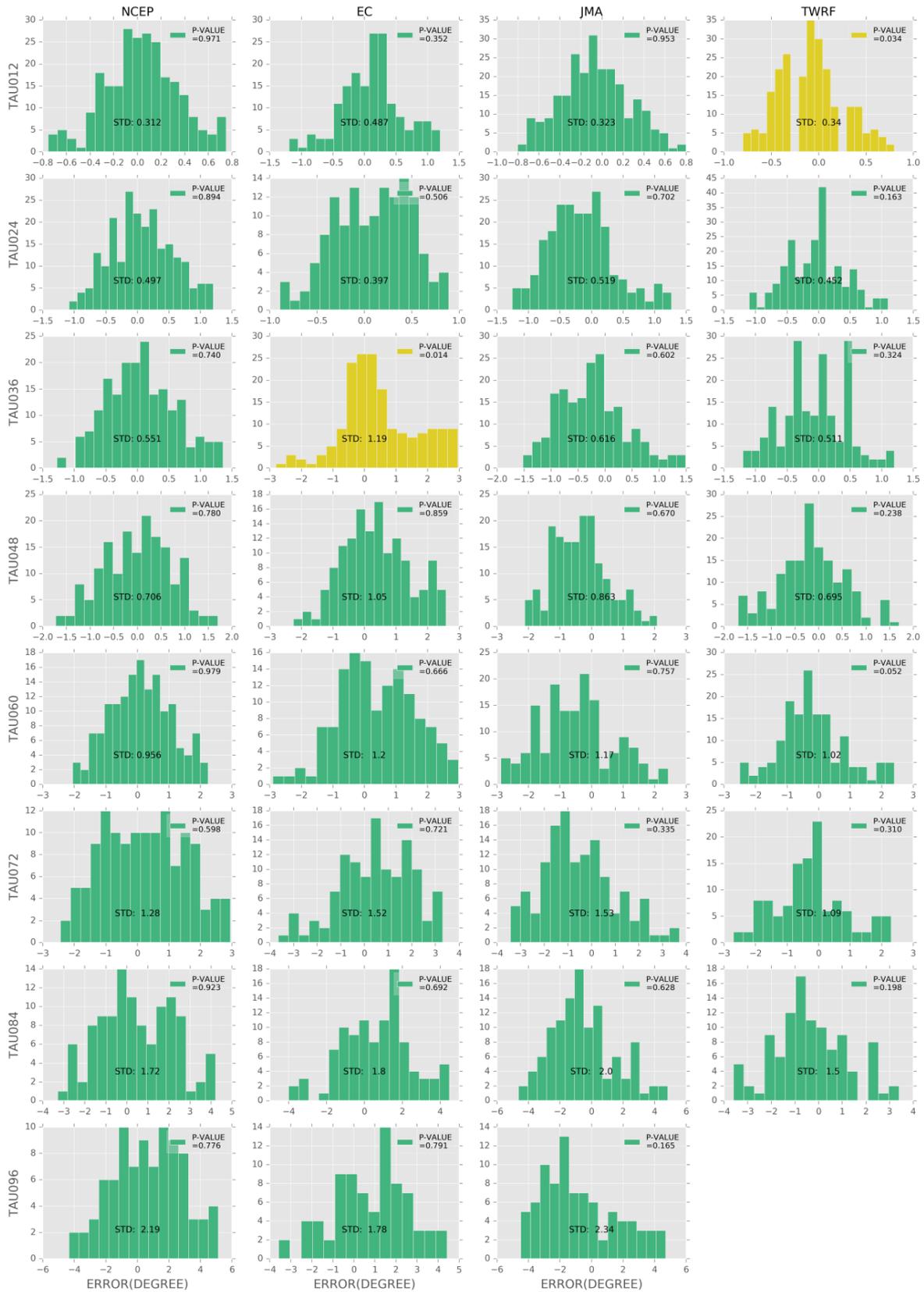


圖 4. 模式在不同預報時間(TAU)的緯度預報誤差分布情形，橫軸單位為度數，縱軸單位為發生次數。在右上角為 Kolmogorov-Smirnov 檢驗該分布是否為常態的 P-Value。柱狀圖的

顏色為綠色代表 P-Value 大於 0.05，小於 0.05 以黃色表示，各圖中央的數字代表標準差。

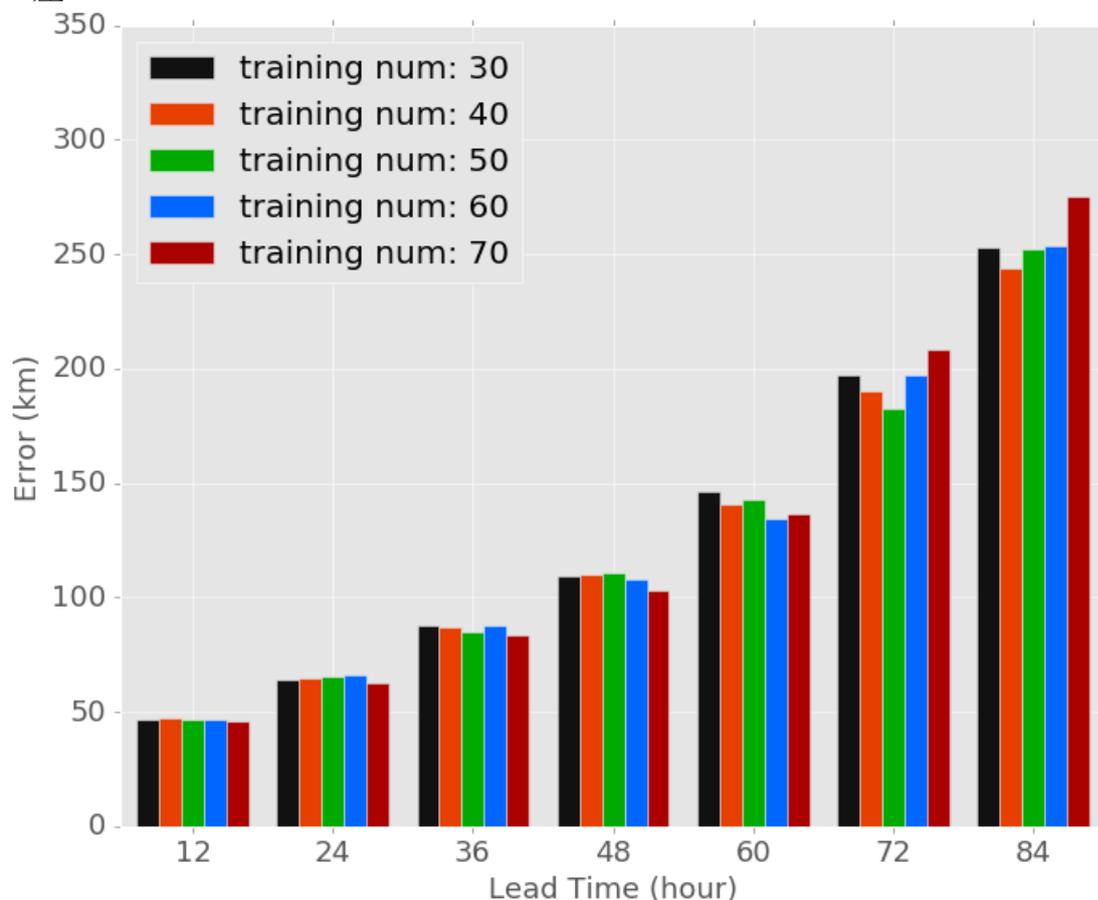


圖 5. 不同訓練期長度對平均誤差的影響測試結果，橫軸為 7 個不同預報時，縱軸代表該次測試的平均誤差值(單位為公里)，不同的顏色代表不同的測試參數，對應表在圖框的左上方。

測試結果如圖 6 所示，橫軸為不同的預報時，不同顏色的直方圖分別代表不同的測試組別，黑色的是未做限制的控制組。圖中可以看出在預報 012~048 小時之間的範圍都沒有明顯變化，060 小時的預報則以控制組的平均誤差最大，而 072 與 084 小時的預報也以控制組的誤差較高。另一方面，072 小時預報的誤差以限制 15 度與 20 度兩組最佳，084 小時的預報誤差則以限制 20 度與 25 度

兩組最佳。空間分布限制所造成的平均誤差差異在 060 小時預報後才逐漸出現可能是因為預報時較長時，可用的訓練資料時間密度相對較低，剔除掉部分超出範圍的資料之後容易找到較早颱風案例資料而增加誤差。基於此測試比較，後續以空間分布限制 20 度為條件選擇訓練期資料。

三、分析結果與討論

(一) 路徑預報誤差分析

圖 7 為 BMA 以及等權重平均的平均路徑誤差值，由上到下分別為經度預報誤差、緯度預報誤差、以地表大圓距離計算出的絕對路徑誤差，橫軸代表不同的預報時，縱軸代表預報誤差，單位上方兩張單位為度，第三張為公里，圖中黑色的直方代表等權重平均路徑(Equal Weighting Consensus; EWC)的平均路徑誤差值，而紅色的則代表 BMA 的平均路徑誤差值，每個預報時上方的百分比數字代表以等權重平均的誤差減去 BMA 的誤差，再除以等權重平均的誤差所得來的改善幅度，從圖上可以看到在預報 012~072 小時的區間中，BMA 的預報結果相較等權重平均為佳，但是改善的幅度隨著預報時增加而減少，路徑誤差方面，在預報 012 小時的改善幅度可達 40.2%，但是到預報 048 小時降低到 21.5%，在預報 072 小時則只有 6.3% 的進步幅度，並且於 084 與 096 小時預報出現了等權重平均法表現優於 BMA 的情況，但是在第 108 以及 120 小時預報則是又以 BMA 略佳，顯示 BMA 預報在 096~120 小時的穩定性相對不好。若是將經度預報誤差與緯度預報誤差分開來討論，從圖 7 中可以發現在 048 小時內的預報誤差改善幅度在經度預報與緯度預報的數值接近，也同樣有著改善幅度隨著預報時間增長而緩慢下降的趨勢在，到了 060 小時之後此特徵開始有了變化，經

度預報的改善幅度快速的下降，在 060 與 072 小時的改善幅度與緯度預報的改善幅度相差許多，而 084 到 120 小時的改善百分比則不明顯，在緯度預報的 012 到 072 小時的改善幅度呈現逐步下降的型態，在 084 到 120 小時則變化穩定性差，沒有一致的特徵。圖 8 為誤差箱型圖，圖中的資料是等權重平均的路徑誤差減掉 BMA 的路徑誤差所得到的結果，橫軸代表的是不同的預報時，而括弧內的則是該預報時所使用的資料筆數，圖中的紅色菱形為平均值，縱軸單位為公里。縱軸在 0 以上代表 BMA 較佳，小於 0 則反之。圖中可看出在 012 小時預報的箱型圖上下四分位所框起來的範圍幾乎都大於 0，而平均值所在的位置高於中位數的高度，單就中位數而言改進幅度只有十幾公里，但是平均值則是高於 25 公里，代表部分資料改善幅度高導致平均值大於中位數的大小。整體而言，中位數的改變幅度相較於每個預報時所算出的平均值是較接近 0 的，在 012~072 小時預報的範圍顯示 BMA 平均有所改進但幅度逐漸減少，到了 084 與 096 小時預報甚至增加了誤差，箱型圖上下四分位數所包住的多數範圍以及中位數也落在 0 以下；而在 108 小時預報的平均值與中位數皆大於 0，但是資料的筆數不多而且數值很分散，在 120 小時預報的平均值與中位數幾乎相同，而資料大多數大於 0。

圖 9 是 BMA 預報誤差與等權重平均預

報誤差的誤差分布圖，橫軸是等權重平均的誤差，縱軸是 BMA 的誤差值，每張小圖的右下角有標明預報時，依序由左上向右遞增，而左上角數字則為 T 檢定的 P-Value 值。圖 9-(A)到 9-(D)顯示誤差分布皆偏向圖形右下，表示 BMA 的預報能力在這段預報時表現較好並且 P-Value 值皆低於 0.05，代表 BMA 模型於 012~048 小時的預報皆具有顯著改善；060~120 小時預報的 P-Value 皆大於 0.05，代表這段預報時區間的 BMA 模型與等權重平均的預報差異並不顯著，而在 084 小時與 096 小時的預報確實呈現等權重平均的預報效果略佳，與前述的圖 7 與圖 8 相呼應。

從上述的誤差分析可知目前實驗所得之 BMA 的改善幅度整體隨著預報時間的增加而降低，最後變得不穩定，推測此現象與單位時間內能得到的訓練資料量，以及成員之間的預報表現差異性有所關聯。假設模式成員在不同時間的表現各有優劣，可能是背景大氣環境的變化，而模式對掌握不同背景場的能力不同，也可能是模式版本調整所致。若某個月內的模式優劣表現都一致，但是在上個月則是相反，那 40 筆訓練資料在同個月內或者是需要兩個月才能湊齊，則預報成效就會因模式參數估計結果而有所差異。假設今天有個颱風從某日的 00:00 UTC 出生共生存了 5 天，每天的 00:00 與 12:00 UTC 會發布一次預報，並且在這 5 天內也都有預報發布，則可以得到能與觀測配對的預報時 024

小時預報有 9 筆，072 小時的則是 5 筆，若該月有 5 個這樣的颱風生成，這樣該月的最後幾個颱風就能在同個月內湊足預報時 024 小時所需的 40 筆資料，但是預報時 072 小時的則需要抓到上個月甚至更早的資料，預報時 120 小時就需要更久以前的資訊。另一點造成改善幅度隨著預報時間增加而下降的因素，推測是模式之間的預報表現優劣差異隨著預報時間的增加而減少。參考圖 3 與圖 4 中各預報時間標示的模式標準差，以經度預報的圖 3 為例，以預報時間 024 與 072 當作前後預報時間的比較基準，預報時間 024 的四個模式由左到右的誤差標準差分別是 0.441、0.504、0.6、0.486，將其除以 0.441 標準化後會得到 1、1.143、1.361、1.102，而預報時間 072 的四個模式誤差標準差為 1.49、1.62、1.83、1.69，分別除以 1.49 之後為 1、1.08、1.23、1.13，可以看到四個模式於誤差標準差的差異縮小。另一方面，圖 4 中緯度預報的預報時間 024 標準差 0.497、0.397、0.519、0.452，標準化後為 1、0.799、1.044、0.909，預報時間 072 為 1.28、1.52、1.53、1.09，標準化後為 1、1.188、1.195、0.852，預報時間 024 與 072 間的模式間差異性沒有明顯變小。當各個成員之間的誤差表現相近，不論是優劣，只要成員的表現無法展現出優劣差異性，調整成員權重的整合方式將難以與等權重平均產生明顯差異，單就預報時間 024 與 072 的比較上可以在經度預報的標準

差差異上看到模式表現差異隨著預報時間而縮減的情形，可能可以解釋預報改善幅度隨預報時間降低的部分原因。

此外，若將不同預報時間的各次預報中，以經度/緯度預報最大值減去最小值，再取出

中位數，做為另一個判定預報成員間差異性的依據，參考表 1 可知模式在整體上緯度預報上的差異大於經度預報，因此我們推斷 BMA 在緯度預報上的改善幅度大於經度預報可能是此差異所造成。

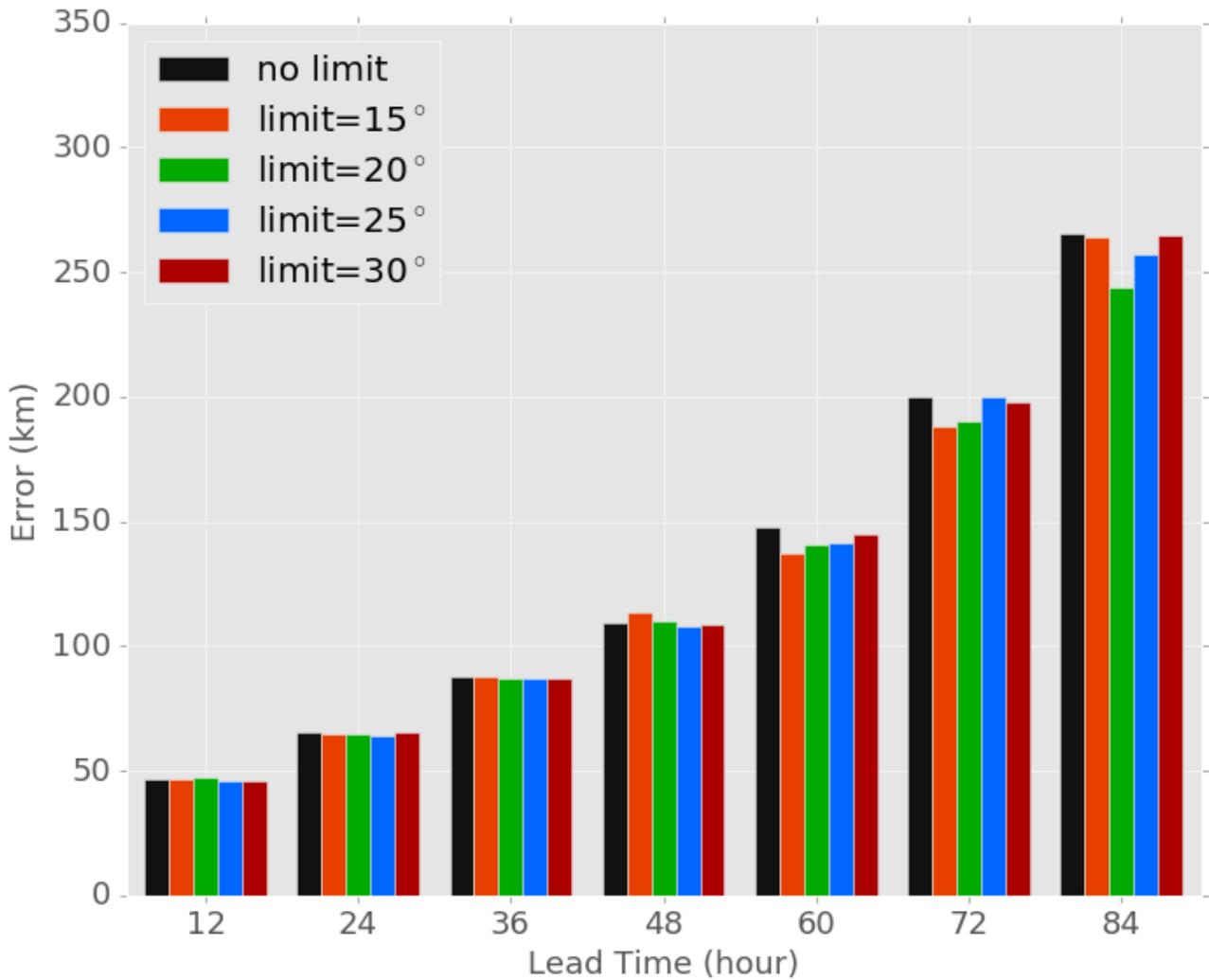


圖 6. 不同訓練期資料空間限制對平均誤差的影響測試結果，橫軸為 7 個不同預報時，縱軸代表該次測試的平均誤差值(單位為公里)，不同的顏色代表不同的測試參數，對應表在圖框的左上方。該度數代表訓練資料在預報時間 000 時的位置必須在該次預報時間 000 位置的經緯度加減 N 度的範圍內。

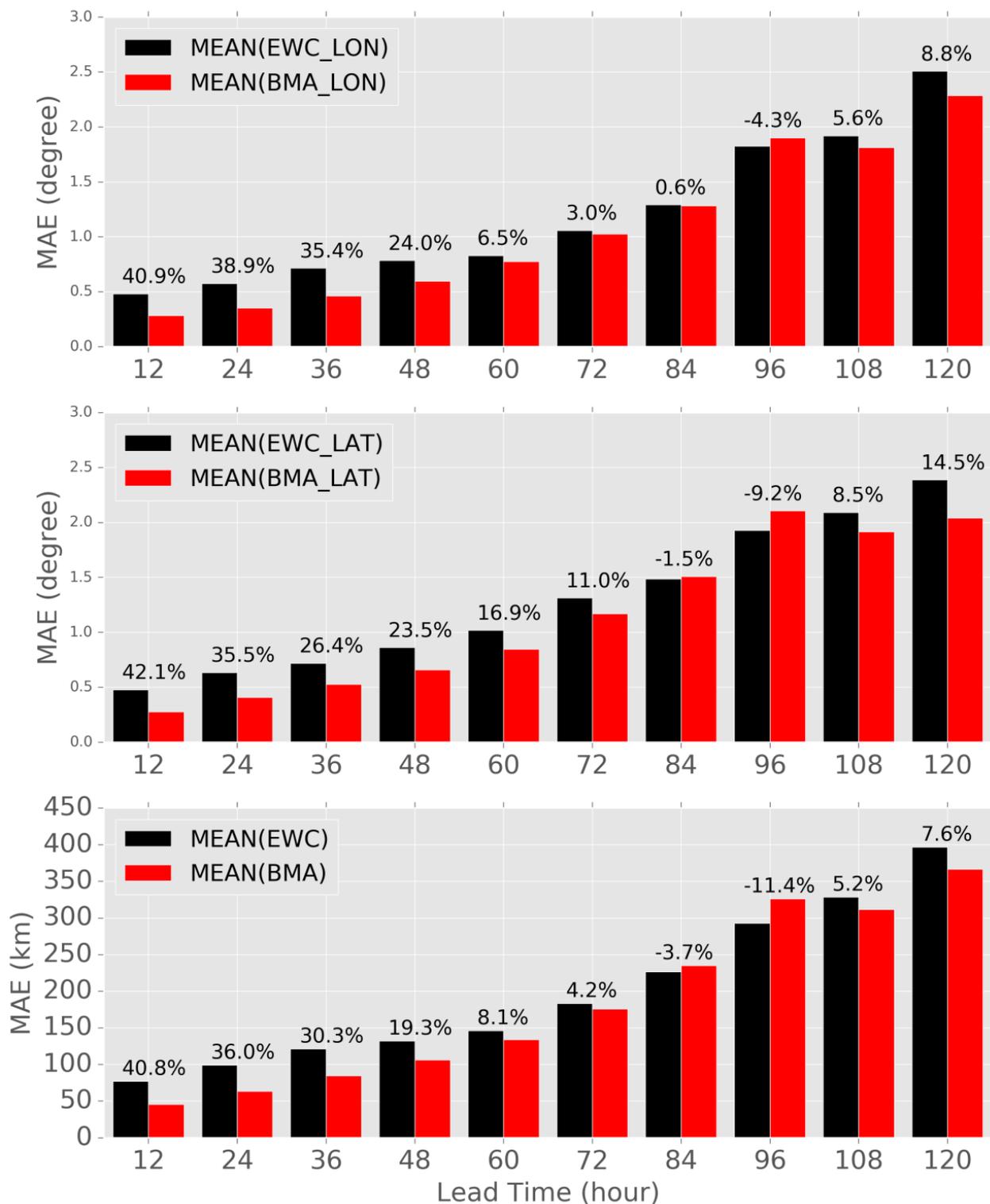


圖 7. BMA 與等權重平均在平均誤差上的表現，由上而下是經度預報誤差，緯度預報誤差，大圓距離誤差，橫軸為 10 個不同的預報時，縱軸為誤差距離(上兩圖單位為度，下圖單位為公里)，黑色代表等權重平均(Equal Weighting Consensus; EWC)路徑的平均誤差，紅色則是 BMA 的路徑預報平均誤差，在各組直方上方所標示的百分比數字代表 BMA 相對於等權重平均的改善幅度。

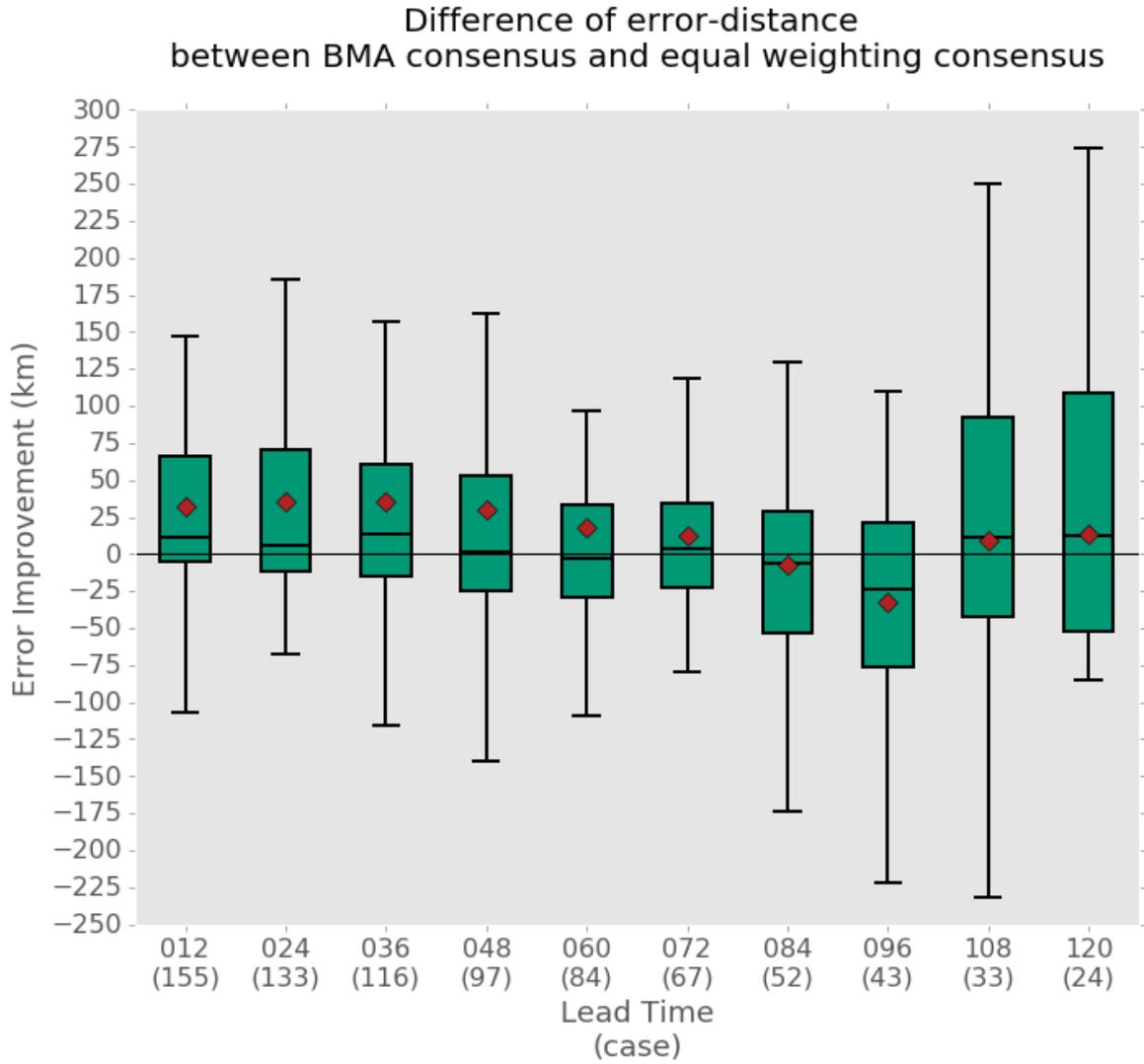


圖 8. 將等權重路徑平均的預報誤差減去 BMA 路徑誤差的箱型圖，橫軸數字代表預報時，括弧內的數字代表資料筆數，縱軸表示 BMA 的改進程度，單位為公里，紅色的菱形代表平均值。

表 1. 2016 年颱風個案中以 NCEP、EC、JMA 與 TWRf 模式的颱風中心位置預報最大值減去最小值的差異中位數

預報時間 (小時)	012	024	036	048	060	072
經度預報 差異 中位數(°)	1.036	1.210	1.234	1.435	1.400	1.590
緯度預報 差異 中位數(°)	1.152	1.541	1.684	2.114	2.076	2.451

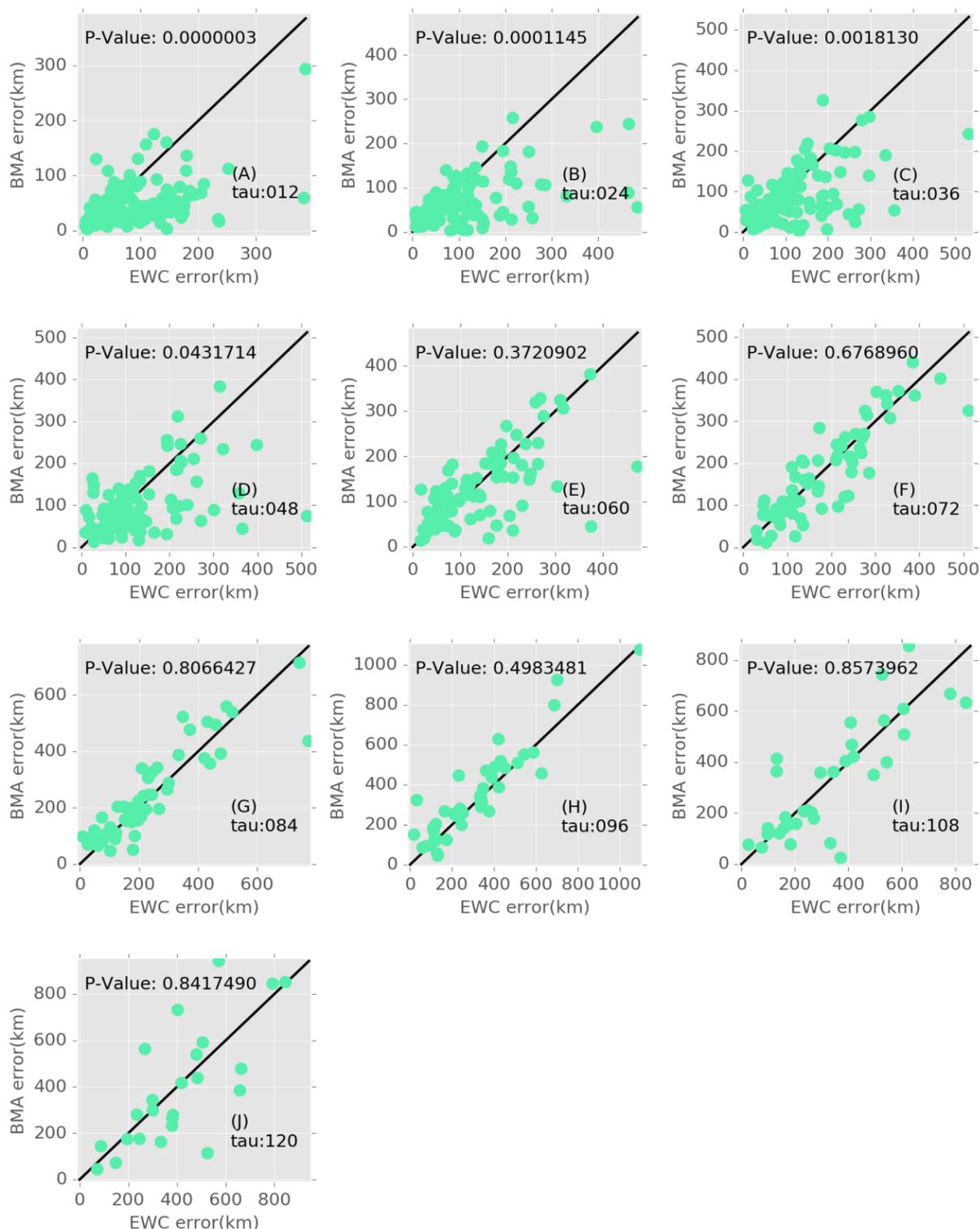


圖 9. BMA 與等權重預報誤差分布圖，橫軸代表等權重平均的誤差，縱軸則是 BMA 的誤差值，每張圖代表的預報時間由左而右依序遞增，並且標示在每張圖右下角之中，而每張圖左上角的 P-Value 是由 T 檢定兩者是否具有相同平均值的結果。

(二) 路徑機率預報能力分析

為了解 BMA 颱風路徑機率預報能力，選用 PIT (probability integral transform) 直方圖以及 CRPS(continuous ranked probability score)技術得分進行分析。假設由 BMA 颱風路徑機率分布所得的累積機率分布 (Cumulative Distribution Function, CDF) 為 $P(x)$ ，則將觀測值帶入 $P(x)$ 後可以得到該次預報的 PIT 值，將各次預報的 PIT 值繪製累積機率與相對頻率即為 PIT 直方圖。PIT 直方圖越接近均勻分布(Uniform Distribution)，則代表路徑機率預報越是準確。假設將累積機率從 0 到 1 的區間分成 10 個區間，亦即各區間的發生頻率為 0.1，若進行 1000 次的預報，理想的 PIT 圖就是各區間的頻率值都為 0.1，也就是說觀測帶入預報的 CDF 函式 $P(x)$ 之後，誤差的分布情形與我們所估計的誤差分布具有一致性。進一步將理想的發生頻率以 0.1 進行標準化，得到相對頻率，亦即若某個區間的數值低於 1，則表示預報高估該區間的發生機率，反之則是低估。另一方面，CRPS 的計算如式(7)所示，其中 $ncase$ 代表校驗資料的筆數， P 代表 BMA 機率預報的分布， O 表示觀測的機率分布，此分布在大於等於觀測值的區間為 1，小於觀測值的區間為 0，CRPS 是將布賴爾評分(Brier Score)對所有門檻值積分並且將所有的案例取平均得到量化數據，當機率報的分布越分散，或者集中在離觀測發生的位置較遠，則

CRPS 值會隨之上升，反之當機率預報的機率分布越集中並且也越準確時，其值越低，代表機率預報的表現越好。(Hersbach, 2000)

$$CRPS = \frac{1}{ncase} \sum_{ncase} \int_{-\infty}^{\infty} [P(x) - O(x)]^2 dx,$$

$$O(x) = \begin{cases} 0, & x < obs \\ 1, & x \geq obs \end{cases} \quad (7)$$

圖 10 是 BMA 颱風路徑經度預報機率預報的 PIT 直方圖，由左至右、由上到下依序呈現 012~120 小時預報的分析結果並於其上標示 CRPS 計算值。觀察可知，012 小時的機率預報於各區間頻率大致上都落在 1 附近，024~072 小時預報機率於 0.2~0.7 的區間數值偏高，表示預報機率略為低估，而 072~120 小時的預報 PIT 圖明顯發現預報能力不確定。圖 11 為等權重平均颱風路徑經度預報機率預報的 PIT 直方圖，在 BMA 的 EM 演算法中，若某個成員在每一筆訓練資料所得到的得分比重都是相同的，則估計出來的誤差標準差會恰好等於直接計算其誤差標準差，也就是說以本文中四個預報成員的例子來說，算式(6)的 \hat{z}_{kt} 中每一個值因為是等權重所以都會是 0.25，將其帶入算式(6)之後，得到的值會與直接計算誤差標準差無異。因此對應的等權重平均採用訓練期資料的誤差標準差作為各模式條件機率常態分布的標準差，整體而言可看出各預報時的 PIT 圖在中央的區間相對發生頻率較高，相反地在左右兩側區間則是高估發生頻率，推斷是因為在各個模

式的預報能力各有優劣的情形之下，等權重平均的方式並未將較多的權重給予較準確的模式亦未縮小誤差較大者的權重所致。將兩者的 CRPS 相比較，可以看到 BMA 的 CRPS 在 012~096 小時的預報皆小於等權重平均者，但是在 108~120 小時則是大於等權重平均者，表示 BMA 颱風路徑經度預報的機率預報能力隨著預報時間的增加而逐漸下降，與 PIT 圖的特徵相呼應。

圖 12 為 BMA 颱風路徑緯度預報的機率預報分析結果，顯示在 012~036 小時的預報表現較佳，在 048 小時與 060 小時預報則於 0.9~1.0 區間出現低估而於 0.1~0.2 區間出現高估，其後機率預報表現則變得不穩定。圖 13 為等權重平均颱風路徑緯度預報的機率預報 PIT 直方圖，呈現與圖 11 相似的表象。逐預報比較可以看出 BMA 的 CRPS 均小於等權重平均，差值大於兩者在經度上的差異，顯示 BMA 颱風路徑緯度預報的機率預報改善幅度優於經度預報。此現象可能是因為颱

風路徑的模式預報能力差異多來自於緯度預報，也就是說模式之間的預報誤差標準差在緯度預報的差異是大於經度預報的，以圖 2 與圖 3 的 048 時預報為例，在經度預報上四個模式的標準差分別為 0.821、0.872、0.906、0.756，而在緯度預報中四個模式的標準差則為 0.706、1.05、0.863、0.695，可以發現在緯度預報中各個模式間的標準差較為分散，而經度預報則相對集中，假設有兩個期望值相同的機率預報，CRPS 計算會給予機率較集中的預報較低的數值，也就是判斷為較佳的機率預報，當透過客觀方法給與模式權重係數並且準確地給予誤差較小的成員時，颱風路徑的緯度預報相較於經度預報將可在 CRPS 獲得較顯著的改善。若今天選用的成員在緯度預報的預報能力表現相近，則 BMA 與等權重平均所做出的 CRPS 差異則會縮小，能夠在這類統計指標上看出改進，必須是成員間本身就有一定程度的差異並且能藉由整合方法給予各成員適合的權重。

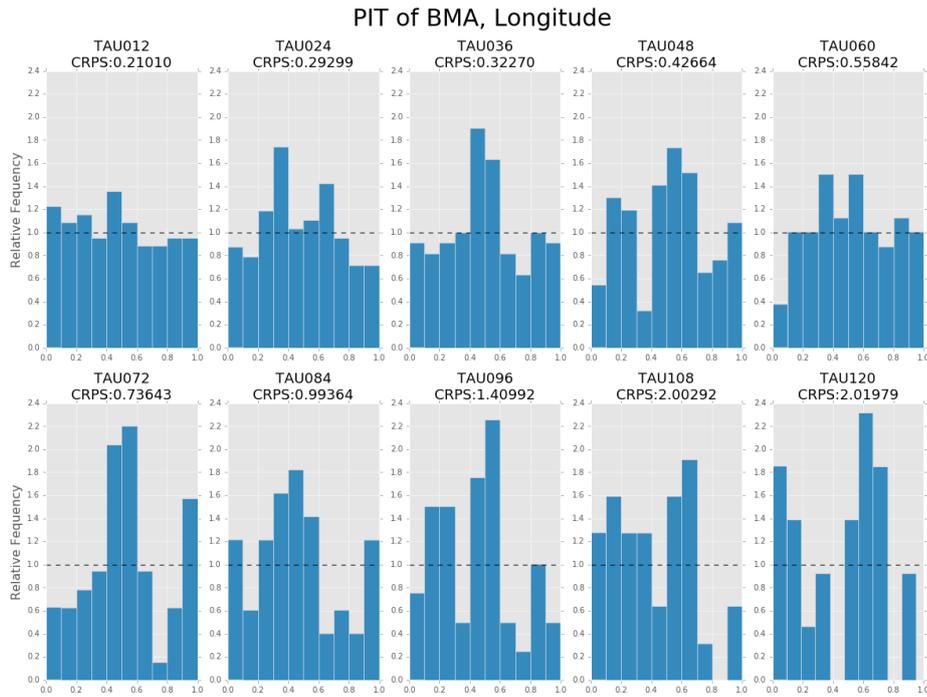


圖 10. BMA 預報結果在經度預報上的 PIT(probability integral transform)直方圖，橫軸代表預測 CDF 值，每個區間的數值大小代表觀測落在 CDF 該區間的頻率，虛線代表理想情況的頻率值。每張圖框上方所標示的是預報時以及 CRPS(continuous ranked probability score)得分。

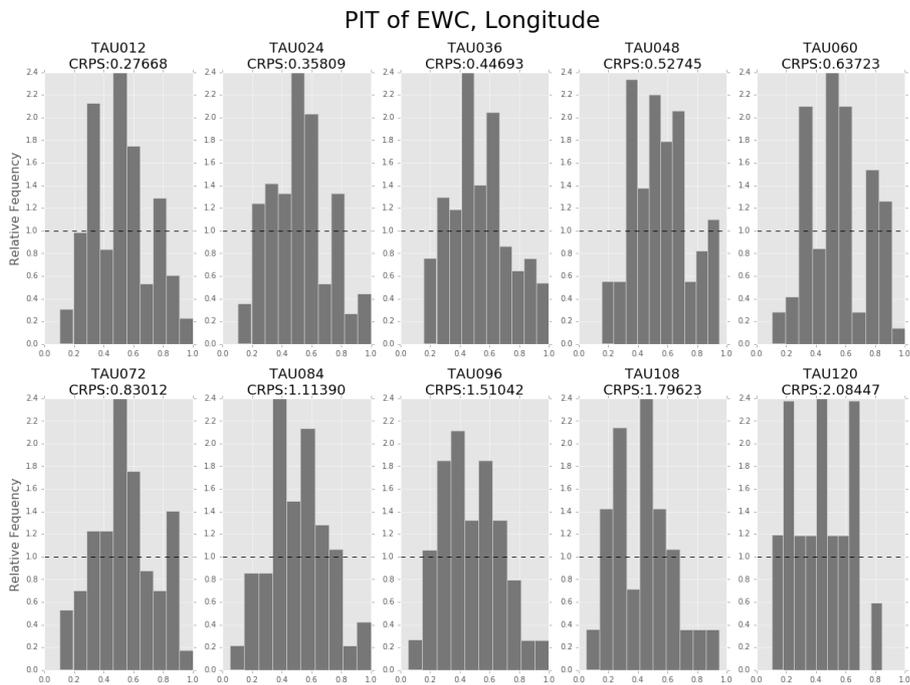


圖 11. 等權重平均預報結果在經度預報上的 PIT(probability integral transform)直方圖，橫軸代表預測 CDF 值，每個區間的數值大小代表觀測落在 CDF 該區間的頻率，虛線代表理想情況的頻率值。每張圖框上方所標示的是預報時以及 CRPS(continuous ranked probability score)得分。

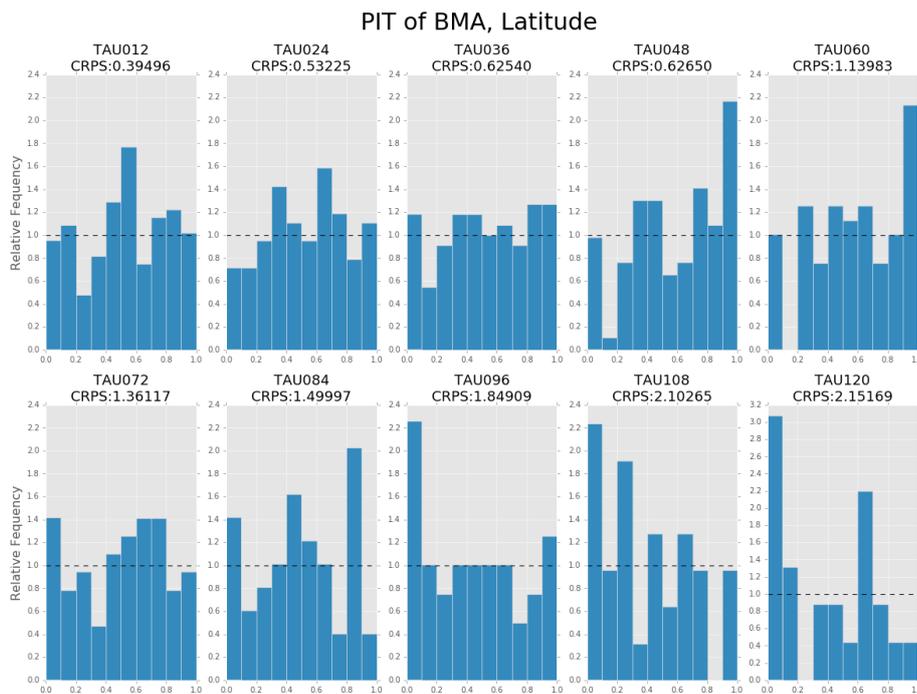


圖 12. BMA 預報結果在緯度預報的 PIT(probability integral transform)直方圖，橫軸代表預測 CDF 值，每個區間的數值大小代表觀測落在 CDF 該區間的頻率，虛線代表理想情況的頻率值。每張圖框上方所標示的是預報時以及 CRPS(continuous ranked probability score) 得分。

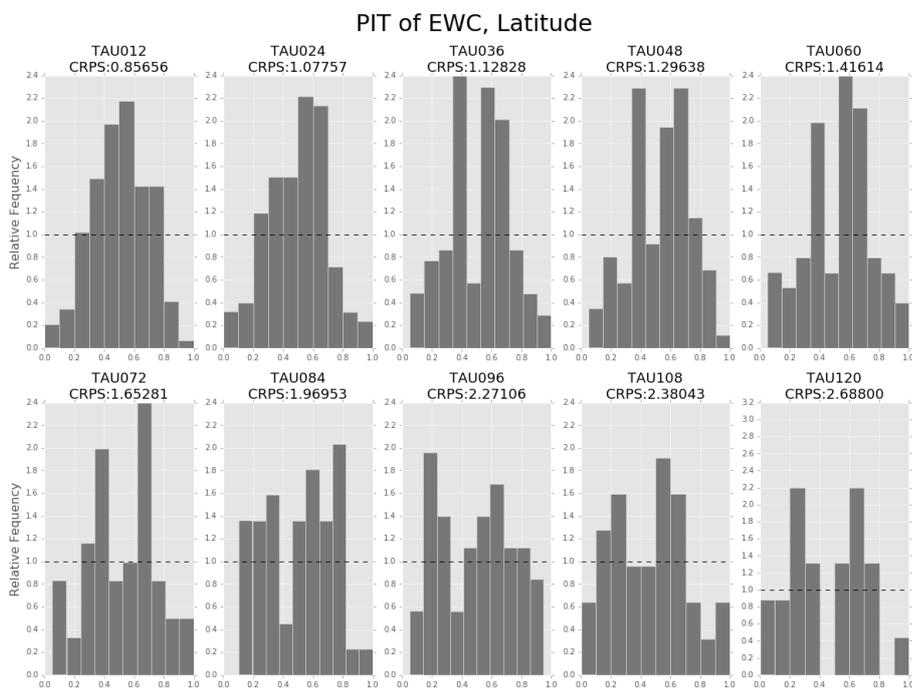


圖 13. 等權重平均預報結果在緯度預報的 PIT(probability integral transform)直方圖，橫軸代表預測 CDF 值，每個區間的數值大小代表觀測落在 CDF 該區間的頻率，虛線代表理想情況的頻率值。每張圖框上方所標示的是預報時以及 CRPS(continuous ranked probability score) 得分。

四、結論

BMA 的優勢在於可利用過去的預報與觀測資料，給予能力較佳的模式較高的權重係數，故當四個模式的中心位置預報差異越大，並且確實存在著優劣之分，越可以得到有別於 EWC 的結果。

本研究應用貝氏模型平均法整合 NCEP、ECMWF、JMA 與中央氣象局 TWRP 模式的颱風路徑決定性預報，並以 2016 年颱風案例進行校驗分析，歸納以下結論。

1. 模式颱風路徑在經度預報與緯度預報的預報誤差分布大多符合常態分布
2. 採用 40 筆以上的訓練期資料建置 BMA 颱風路徑模型可得到穩定預報成效，並且限制訓練期資料的空間分布對於短時間內的預報成效影響較小，在 060 小時以後的預報才有明顯影響。
3. BMA 颱風路徑期望值在 012~048 小時以內的預報顯著優於等權重平均，24 與 48 小時的預報誤差可分別降低至約 65 公里(改善 36%)與 110 公里(改善 19.3%)，而 060 小時以後的預報則無顯著差異。
4. BMA 颱風路徑機率預報較等權重平均者為佳，並且於緯度預報的預報改善幅度較大。
5. 當四個模式的中心位置預報差異越大，BMA 緯度預報有較佳的優勢掌握真實路徑，但 072 小時以後的預報表現較不穩定。

實作貝氏模型平均法整合多模式的颱風路徑決定性預報得到颱風路徑於經度預報與

緯度預報的機率密度函數，並以期望值作為颱風中心位置的決定性預報，確實有助於提升颱風路徑預報能力並發展機率預報指引。以現階段的實驗結果而言，我們認為 BMA 方法與其他整合方式比較起來最明確的優勢是可以藉由條件機率與誤差分布來得到機率型預報，預報屬於混沌系統的大氣變化，掌握誤差的分布情形在小至日常生活大至災害防治的應用上，有時候是更勝於單純的決定性預報。BMA 誤差分布的估計來自於過去的預報表現，若是模式長期表現都沒有變化，在各區域各季節的表現也一致，則隨意選擇足量筆數的訓練資料即可，目前最顯而易見的限制就是選取的訓練資料必須要對當下的颱風預報有指標性，若是預報模式在不同年份不同季節不同區域的誤差表現有所不同，甚至是分布型態改變，則訓練之後得到的權重反而會增加誤差量。未來可進一步朝訓練期資料選擇方式以及如何提升 060 小時以後的 BMA 模型預報能力進行測試與改進。

致謝

本研究在中央氣象局氣象預報中心委辦計畫「105 年度小區域暨災害性天氣資料分析與應用建置案-統計降尺度天氣預報子系統」支援下完成。

參考文獻

1. 蔡孝忠、呂國臣、許乃寧、賈愛玫，2011：蒙地卡羅法在颱風侵襲機率估計的應用。大氣科學期刊，39(3)，269-288

2. 張語軒、張庭槐與吳蕙如，2011：貝式模型平均於溫度機率預報之應用。氣象學報，49(1)，19-38。
3. 馮智勇、劉家豪、許乃寧、賈愛玫、黃嘉美，2015：應用貝氏模型平均法發展臺灣測站氣溫機率預報指引。104 年天氣分析與預報研討會論文彙編，中央氣象局，臺北。
4. DeMaria, M., J. A. Knaff, R. Knabb, C. Lauer, C. R. Sampson, and R. T. DeMaria, 2009: A new method for estimating tropical cyclone wind speed probabilities. *Wea. Forecasting*, 24, 1573–1591
5. Dempster, A. P., N. M. Laird, and D. B. Rubin, 1977: Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Stat. Soc.*, 39B, 1–39.
6. Dong, L., and F. Zhang, 2016: OBEST: An observation-based ensemble subsetting technique for tropical cyclone track prediction. *Wea. Forecasting*, 31, 57–70, doi:10.1175/WAF-D-15-0056.1.
7. Elsberry, R. L., 2014: Advances in research and forecasting of tropical cyclones from 1963–2013. *Asia-Pac. J. Atmos. Sci.*, 50, 3–16.
8. Elsberry, R. L., J. R. Hughes, and M. A. Boothe, 2008: Weighted position and motion vector consensus of tropical cyclone track prediction in the western North Pacific. *Mon. Wea. Rev.*, 136, 2478–2487.
9. Fisher RA. 1922 On the mathematical foundations of theoretical statistics. *Phil. Trans. R. Soc. Lond. A* 222, 309–368. (doi:10.1098/rsta.1922.0009)
10. Hersbach, H., 2000: Decomposition of the continuous ranked probability score for ensemble prediction systems. *Wea. Forecasting*, 15, 559–570.
11. 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.]
12. Hsiao, L.-F., C.-S. Liou, T.-C. Yeh, Y. R. Guo, D.-S. Chen, K.-N. Huang, C.-T. Terng, J.-H. Chen, 2010: A vortex relocation scheme for tropical cyclone initialization in advanced research WRF. *Mon. Wea. Rev.*, 138, 3298–3315.
13. Hsiao, L.-F., D.-S. Chen, Y.-H. Kuo, Y.-R. Guo, T.-C. Yeh, J.-S. Hong, C.-T. Fong, and C.-S. Lee, 2011: Testing of WRF 3DVAR for model initialization and its application on operational typhoon prediction in Taiwan. Paper submitted to *Mon. Wea. Rev*
14. Jun, S., W. Lee, K. Kang, K. Y. Byun, J. Kim, and W. T. Yun, 2015: Applicability of the superensemble to the tropical cyclone track forecasts in the western North Pacific. *Asia-Pac. J. Atmos. Sci.*, 51,

39-48.

15. Kass, R. E., and A. E. Raftery, 1995: Bayes factors. *J. Amer. Stat. Assoc.*, 90, 773–795.
16. Krishnamurti, T. N., C. M. Kishtawal, Z. Zhang, T. LaRow, D. Bachiochi, E. Williford, S. Gadgil., and S. Surendran, 2000: Multimodel ensemble forecasts for weather and seasonal climate. *J. Climate*, 13, 4196-4216.
17. Leamer, E. E., 1978: *Specification Searches*. Wiley, 370 pp.
18. McLachlan, G. J., and T. Krishnan, 1997: *The EM Algorithm and Extensions*. Wiley, 274 pp.
19. 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.
20. Roberts, H. V. , 1965: Probabilistic prediction. *J. Amer. Stat. Assoc.*, 60, 50-62.
21. Williford, C. E., T. N. Krishnamurti, R. Correa Torres, S. Cocke, Z. Christidis, and T. S. Vijaya Kumar, 2003: Real-time multi-model super-ensemble forecasts of Atlantic tropical systems of 1999. *Mon. Wea. Rev.*, 131, 1878-1894.

Using Bayesian Model Averaging to Develop Probabilistic Guidance for Tropical Cyclone Track Forecast

Yu-Tsung Chen¹ Chih-Yung Feng¹ Po-Hsiung Chang²
Nai-Ning Hsu² Aimei Chia²

¹Manysplendid Infotech, Ltd.

²Central Weather Bureau

(manuscript received 17 October 2017 ; in final form 29 August 2018)

Abstract

With the improvement of numerical weather prediction models and the raise of multi-model and ensemble forecasts concept, the average forecast error of tropical cyclone track is continuously decreasing, which makes specifying the track forecast uncertainty much more realistic. This study proposes using Bayesian Model Averaging (BMA) for statistically integrate multi-model track forecasts along longitude and latitude separately. The BMA predictive probability density function (PDF) of longitudinal or latitudinal position is a weighted average of PDFs centered on the individual forecasts, where the weights are estimated from models' predictive performance over the training period. Consequently, the BMA deterministic forecast of typhoon center can be obtained from the expectation of PDF.

This method was applied to 2016 typhoon cases using NCEP (National Centers for Environmental Prediction), ECMWF (European Centre for Medium Range Weather Forecasts), JMA (Japan Meteorological Agency) and CWB (Central Weather Bureau) Typhoon WRF models. Compare to the common equal-weighting consensus forecast, the BMA expectation is better at 12-h, 24-h, 36-h and 48-h average forecast error according to 95% confidence interval T test. The 24-h and 48-h forecast error of typhoon position are 65 km and 110 km respectively. Furthermore, the improvement in BMA latitudinal position forecast is larger than the longitudinal one because of the larger latitudinal deviation of model forecasts availing BMA against distinct weightings according to model performances. Consequently, the BMA PDF can distinguish the most likely typhoon position from the others and be a powerful consensus probabilistic guidance for the CWB typhoon forecasts operation.

Keywords: Probabilistic Typhoon Track Guidance, Bayesian Model Averaging. doi: 10.3966/025400022018064602003