

臺灣地區 WRF 颱風系集降雨機率預報之評估、校正 與經濟價值分析-第一部分：預報評估

張惠玲¹ 陳冠儒¹ 吳佳蓉¹ 汪琮¹ 洪景山¹ 楊舒芝²

¹中央氣象局

²國立中央大學大氣物理研究所

(中華民國 107 年 3 月 5 日收稿，中華民國 107 年 5 月 04 日定稿)

摘要

本研究主要是針對臺灣地區的颱風，利用中央氣象局發展的WRF系集預報系統(WEPS)，進行定量降雨機率預報；最終目的在於提供更具參考價值的系集颱風降雨預報產品，並讓使用者藉由系集機率預報做出最佳化的決策，以得到最大的經濟價值。研究成果擬分述於三篇論文中，包括系集降雨機率預報之評估、校正與經濟價值分析。

本論文(第一部分-預報評估)的研究重點在於評估WEPS的品質良窳，並利用WEPS產生颱風降雨機率預報，詳細校驗其預報表現。研究結果顯示：WEPS的系集散度可良好地反應預報的不確定性。颱風降雨機率預報具有相當好的區辨能力，但有明顯的預報偏差，可信度和區辨能力均隨著降雨門檻增加而降低。

不同區域的校驗結果顯示：颱風降雨預報在陸地區域的可信度與區辨能力均優於海洋區域，這與做為真實場的雷達降雨估計特性有關。此外，平地的可信度優於山區，但山區的區辨能力優於平地。整體而言，山區的預報技術優於平地，這是因為地形鎖定效應造成颱風降雨在迎風斜坡上的可預報度高於平地，因而WEPS在山區的預報技術和區辨能力相對較佳。Mann-Whitney檢定顯示山區的預報能力優於平地具有統計上的顯著性。

關鍵詞：系集降雨機率預報、可信度、區辨能力、預報技術

一、前言

颱風所伴隨的豪(大)雨常常重創臺灣地區，造成經濟損失，並威脅人民的生命安全和財產，因而颱風定量降雨預報深受政府防災單位和民眾所重視。近年來，隨著系集預報的發展，定量降雨預報得以往前推展到定量降雨機率預報，讓民眾不單獲得降雨量的預報資訊，更進一步得知預報的不確定性。

不同於傳統單一預報的決定論觀點，系集預報可以根據有限的系集成員對未來的天氣狀態產生近似的機率分布，進一步提供使用者預報的可信度資訊(Toth and Kalnay 1993)。關於系集預報系統的組成設計，最重要的是能夠攫取與預報誤差相關的各種不確定性來源，諸如來自觀測資料、模式初始場、資料同化過程和數值模式的不確定性。針對如何產生系集初始擾動，過去研究已提出各種不同的方法，諸如 singular vector (Hartmann et al. 1995；Molteni et al. 1996；Buizza et al. 2005)、time-lagged averaging (Dalcher et al. 1988)、breeding method (Toth and Kalnay 1993；1997)、ensemble transform technique (Wei et al. 2008)以及透過系集資料同化系統產生初始

擾動 (Buizza et al. 2008)。除了初始場的不確定性外，另有其它系集擾動方法來獲得模式的不確定性，諸如採用不同的參數化方法、不同的動力模式或對模式進行隨機擾動等等 (Fujita et al. 2007；Lang et al. 2012；Hsiao et al. 2013)。

針對系集預報之評估，我們可以透過各種定量校驗方法來瞭解系集預報系統 (ensemble prediction system, EPS) 在不同方面的預報品質，諸如正確性、可信度、區辨能力、預報技術等等 (Murphy 1993)，常見的校驗方法包括：spread-skill relationship (Kalnay and Dalcher 1987；Whitaker and Lough 1998；Zhu 2005；Buizza et al. 2005；Chang et al. 2012)、rank histogram (RH；Hamill 2001)、reliability diagram (Hsu and Murphy 1986；Hamill 1997)、relative operating characteristic (ROC；Mason and Graham 1999；Jolliffe and Stephenson 2003；Hamill and Juras 2006)、Brier skill score (BSS) 以及 rank probability skill score (RPSS；Wilks 2011) 等等。

幾乎所有的預報系統都存在系統性偏差；過去的研究指出：針對有偏差之預報系統，後端校正是相當重要的程序，可以修正偏差，並提升預報能力 (Mass 2003)。機率預

報的校正方法包括模式輸出統計 (Model Output Statistics, MOS; Glahn and Lowry 1972)、類神經網路法 (Artificial Neural Network techniques, ANN; Mullen and Buizza 2004; Yuan et al. 2007) 以及線性迴歸法 (Linear Regression method, LR; Lu et al. 2007; Yuan et al. 2008; Chang et al. 2012) 等, Chu et al. 2007、Kim et al. 2010 及 Norton et al. 2011 都曾採用 LR 以及 ANN 進行颱風預報的校正。過去研究指出校正結果深受訓練樣本數是否充足 (Atger 2003) 以及樣本間的相依性 (Eckel and Walter 1998) 所影響, 非線性的 ANN 比線性的 LR 需要更多的訓練樣本才能建立穩定的校正關係。此外, 利用具有相似氣候特徵之長期資料做為訓練樣本(例如依地形或氣候特性進行分區校正)可以有效改善校正結果 (Yuan et al. 2008)

Chang et al. (2015) 利用局地分析與預報系統 (Local Analysis and Prediction System, LAPS) 探討臺灣地區地形對於颱風降雨預報能力之影響, 指出中央山脈在颱風降雨之空間分布上扮演相當重要的角色。亦即, 颱風降雨的空間分布主要決定於颱風中心和中央山脈的相對位置, 此即所謂的地形鎖定效應 (terrain-locking effect; Su et al. 2012)。由於地

形鎖定效應, 颱風引發之大降雨經常發生在迎風面的斜坡上(地形高度 ≥ 500 m); 因此, 在迎風斜坡上, 颱風降雨的可預報度通常高於平地, 且 LAPS 降雨機率預報的預報技術和區辨能力也相對較佳。這是一個相當有趣的現象, 因為相較於複雜的山區地形, 數值預報模式通常在平地有較佳的預報表現, 但針對臺灣地區的颱風降雨, 結果恰好相反。

系集機率預報優於單一預報的原因在於考慮了預報過程中的不確定性 (諸如初始誤差、非線性動力誤差和模式誤差等), 並透過機率的方式將不確定性傳達給使用者。當預報機率比較小 (亦即, 所有成員的預報結果比較分歧) 時, 除了代表此天氣事件發生的可能性比較小且預報誤差比較大之外, 同時透露一個訊息: 此天氣事件的不確定性相對較大, 亦即此天氣事件不易準確地被預報。然而, 相較於提供「是」或「否」的單一預報, 機率預報 (不確定資訊) 是進一步幫助使用者在日常生活中做出更好的決策, 或是因無法像單一預報提供「是」或「否」的單一建議, 反而讓使用者難以做決定呢?

事實上, 習慣了傳統單一預報的使用者, 會因為不清楚機率預報

的真正意義而感到困惑。舉例來說，對一個關心氣溫是否低於 0°C (寒害或霜害) 的農夫而言，單一預報提供的資訊是：明日的氣溫是 5°C 。雖然此預報資訊未透露可信度有多大，但農夫會假定此預報完全正確，並依據此預報資訊做決策。然而，機率預報的資訊是：明日氣溫低於 0°C 的機率是80%；不明白機率預報真正涵意的農夫會對80%感到不知所措，到底80%代表「會發生」還是「不會發生」呢？正因如此，大部分的使用者對於機率預報最常提出的要求是：在提供機率預報的同時，能否也提供一個最佳的機率門檻，讓使用者在預報機率超過此門檻時，就視為所預報的天氣事件將會發生，而得以事先採取防災行動，比方說：封路、提早採收作物、搭建遮雨棚或停止上班上課等等。簡言之，使用者最大的困擾是：到底要如何利用機率預報做出最好的決策呢？

最好的決策在本研究中是指：使用者藉由參考預報資訊而得以採取最適當的防災行動，以降低防災成本或減少災害損失，而使長期平均的支出費用(亦即，費用期望值)降到最低。由於不同的使用者在面對同樣的災害性天氣時，會有不同的

致災風險；因此，防災行動必須隨著使用者的不同而有不同的門檻設定。舉個例子來說，在同一地區有種植不同作物的兩位農夫，當降水超過 50 mm (6 h) 時，兩農夫的作物都會受損而必須事先採取防災行動。假定兩農夫採取防災行動的成本是一樣的，但因作物的脆弱程度或價值不同，使得農夫甲採取防災行動後可避免的損失遠小於農夫乙。在這種情況下，兩農夫應該會有不同的防災標準(亦即，兩農夫採取防災行動的最佳機率門檻是不同的)，農夫乙理當比較傾向於採取防災行動。

關於如何利用機率預報做出最好的決策，Murphy (1977)證明如果是採用校正後接近完全可信(亦即沒有預報偏差)的機率預報，那麼在明確知道使用者的成本損失比 $r(r=C/L)$ ， C 是防災行動的成本， L 是採取防災行動後可避免的損失)的狀況下，只要選擇和 r 值相同的機率門檻，就可得到最大的經濟價值。但在某些情況下，使用者的 r 並無法明確被得知，例如Chang et al. (2015)中的應用實例：紅棗易受颱風豪雨的影響而導致果實裂開或品質受損，因此農民必須決定是否趕在颱風來襲之前，提早採收紅棗以減少損失。相較於「正常採收」，「提早採收」看似不需

要任何防災成本，但此行動確實造成農民總收入的減少，因為尚未完全成熟的紅棗價格是比較低的。假設紅棗提早採收的價格只有成熟價格的 R_1 ，而大雨影響後的紅棗價格會下跌至原價的 R_2 。在這情況下，農民該如何做出最好的決策呢？事實上，在這樣的例子中，使用者依然可以透過經濟價值分析，藉由過去的防災經驗求出成本損失比，進而做出最好的決策(Chang et al. 2015)。

Chang et al. (2015)同時也透過一些敏感度實驗，詳細探討預報系統的區辨能力(區辨能力是預報品質的重要指標之一)和經濟價值之間一致和相關的特性。重要發現包括：(1) 預報系統的潛在經濟價值(potential economic value, PEV；定義為預報系統能提供給所有使用者的最大經濟價值總和)主要決定於系統的區辨能力。這樣的結果同時暗示著：預報系統所能提供的PEV只能透過改善系統的區辨能力來加以提升。(2) 預報系統的區辨能力或是所能提供的PEV對於預報偏差並不敏感。換言之，我們可以透過校正來改善預報系統的可信度；但無法改善區辨能力以及所能提供的PEV。此外，倘若使用者可以選擇適當的機率門檻(此機率門檻可以經由過去長期的

經濟價值分布而求得)，即便採用有偏差的預報，依然可以得到最大的經濟價值。(3)相較於單一預報，系集機率預報具有較高的區辨能力，且可提供較大的經濟價值給較多的使用者。這樣的優勢隨著降水強度的增加而愈加顯著，說明了系集機率預報的發展對於颱風強降雨事件的預報是相當重要的。

本研究主要是針對臺灣地區的颱風，利用中央氣象局發展的WRF系集預報系統(WRF ensemble prediction system, WEPS)，進行定量降雨機率預報(probabilistic quantitative precipitation forecast, PQPF)，並進一步對系集機率預報進行評估、校正和經濟價值分析，最終目的在於幫助防災決策者及一般民眾在颱風期間能夠藉由參考系集機率預報做出最好的決策，獲得最大的經濟價值，以提升防颱作業成效。所有研究成果擬分述於三篇論文中，包括系集降雨機率預報之評估、校正與經濟價值分析。第一篇論文(本論文)的研究重點在於評估WEPS的品質良窳，並利用WEPS產生颱風降雨機率預報，詳細校驗其預報表現，找出系統性偏差。

相較於Chang et al. (2012) 和Chang et al. (2015)，本研究有三個新

的科學議題探討，包括：(1)針對臺灣地區的颱風降雨機率預報，山區比平地具有較高的可預報度以及較佳的預報能力，是否具有統計上的顯著性？(2)相較於Chang et al. (2012)所採用的LR校正方法，本研究新增的非線性ANN方法，有何優勢或差異？(3)因防災單位之需求或水文動力模式只能以決定性降水預報作為模式初始條件之需要，氣象局將系集降雨預報透過系集平均、PM (probability-matched mean) 或 QPFP (QPF percentile)方法轉成決定性的系集降雨預報。在轉換過程中，究竟損失了哪些系集預報原有的資訊或優勢呢？上述第一個科學議題擬於本論文中探討。

在Chang et al. (2015)中，我們觀察到LAPS系集降雨機率預報在山區優於平地之特別現象，並推論此現象是地形鎖定效應所造成。但無法確定此現象是否具有普遍性，或只是因為LAPS相對簡單的資料同化和特別的參數法設定所造成，同時也未證明山區優於平地的結論是否具有統計上的顯著性。本論文除採用WEPS和不同的颱風個案再次驗證山區預報能力優於平地之結論外，更重要的是透過統計測試 (Mann-Whitney test)來檢驗此降雨預報特性

(山區優於平地)是否具有統計上的顯著性。藉此了解臺灣地區中央山脈對於颱風降雨的地形鎖定效應，究竟對山區和平地的預報能力有多顯著的影響，進而建立預報員對預報結果的信心。

此外，本篇論文最重要的目的在於完整地評估WEPS的預報品質，包括WEPS是否為品質良好的系集預報系統，以及QPFP在不同區域的預報品質差異。這些巨細靡遺的預報評估是後續兩篇論文的基石，因為透過評估得知預報偏差之地域性差異後，將有助於本研究第二部份針對不同偏差特性之區域進行分區校正，同時也助於確認後續兩個科學議題所得到的結論是正確且具有科學意義的。

本論文第二部分主要介紹WEPS、預報校驗的觀測資料以及所採用的颱風個案。第三部分是研究方法，包括系集預報評估方法、拔靴法與Mann-Whitney檢定。第四部分則是結果與討論，包括系集散度、預報偏差、區辨能力、預報技術之評估結果以及山區和平地的雨量預報表現比較。最後一部分是結論及展望。

二、模式與資料

(一) WRF 系集預報系統 (WEPS)

本研究所採用的 WEPS(Li and Hong 2011, 2014)是使用 WRF/ARW 3.3.1 版，擁有三層巢狀網格，水平解析度分別為 45、15 及 5 公里，模式範圍如圖 1 (陳和洪 2017) 所示。垂直設定共有 45 層，頂層為 30hPa。WEPS 使用 10 組邊界條件擾動、20 組初始場擾動和 20 組不同之模式物理參數法設定，產生 20 組系集成員。第一層巢狀網格之模式初始場來自 CWB WRF M00 模式之前六小時預報場，並加入隨機初始擾動。此隨機擾動是基於 WRF 三維變分分析系統所使用的背景誤差特性所建立。

第二層和第三層巢狀網格之模式初始場由第一層巢狀網格內插。模式所需的邊界條件來自美國國家環境預報中心 (National Centers for Environmental Prediction, NCEP) 全球系集預報系統 (Global Ensemble Forecast Systems, GEFS) 之預報場。預報作業每日執行 4 次，分別為每日的 00、06、12 和 18 UTC。本研究採用其解析度最高的第三層網格資料進行系集颱風降雨機率預報。相較於 Chang et al. (2015) 所建置的 LAPS 系集預報系統，此系集更能代表非線性動力誤差或模式誤差等不確定性。

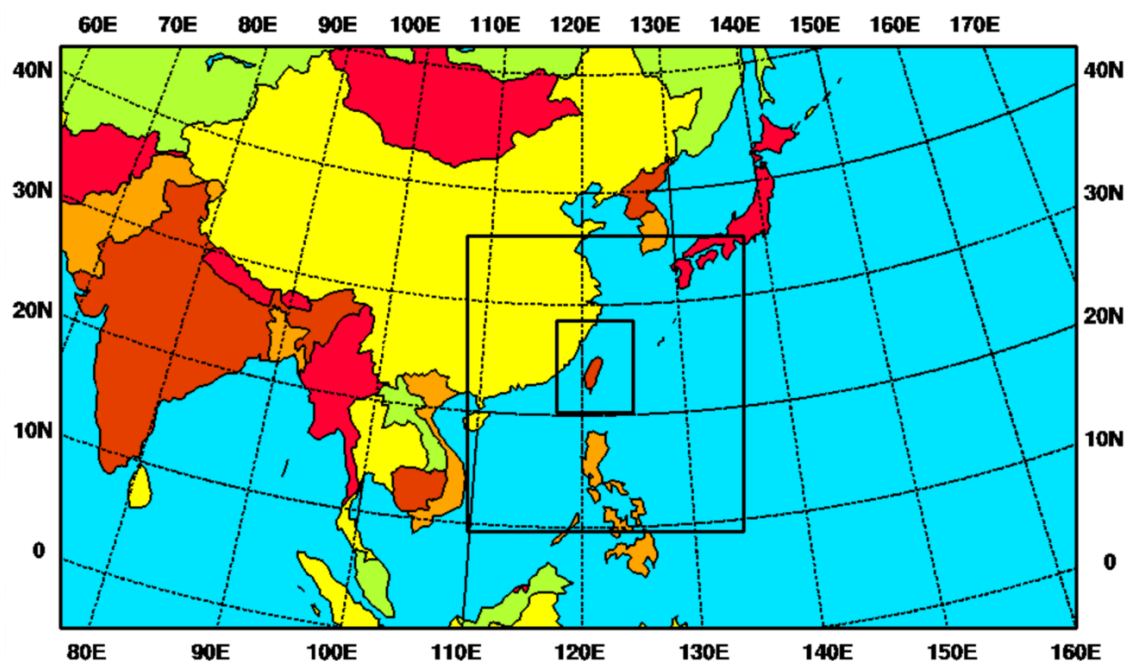


圖 1、CWB WEPS 模式範圍(陳和洪 2017)。

(二)觀測資料

本研究用以進行預報校驗的觀測資料，在降雨部分是使用中央氣象局、經濟部水利署與美國劇烈風暴實驗室 (National Severe Storm Laboratory, NSSL) 所合作發展之劇烈天氣監測系統 (Quantitative Precipitation Estimation and Segregation Using Multiple Sensors, QPESUMS) 的雷達估計降雨 (Quantitative Precipitation Estimation, QPE) 做為觀測資料 (Gourley et al. 2001)，解析度是 $0.0125^\circ \times 0.0125^\circ$ ，其範圍顯示於圖 2 中。要特別說明的是：(1) 陸地區域的雷達 QPE 有經過雨量計的校正，但海洋地區則無法校正，因此陸地區域有較高的正確性；(2) 距離雷達站較遠的海洋地區，雷達波束離地面較高，因此 QPE 低估的可能性相當高；(3) 山區的雨量計密度遠較平地低 (圖 2) 且陸地區域未經校正的雷達 QPE 主要呈現低估現象，由此可推知山區之雷達 QPE 低估的可能性很高。

雷達 QPE 主要是利用雷達最低仰角回波，透過回波反射率 (reflectivity ; Z) 與降雨率 (R) 的關係

(Z-R 關係式) 進行降雨估計。海洋區域的雷達 QPE 低估可能是因為颱風較強的回波區 (>20 dBZ) 主要位於 5 km 高度以下，但中央氣象局作業雷達的最低仰角 (0.5°) 波束大約在距雷達站 220 km 左右的高度就已經高於地表或海面 5 km。換言之，距離雷達站 220~460 km 的回波掃描範圍內，利用最低仰角之回波大小所估算的降雨很有可能低於地面或海面上實際觀測到的降雨。此外，海洋區域的回波觀測距離雷達站較遠，雷達回波經過天氣系統後所造成的能量衰減，也是 QPE 低估的原因之一。氣象局針對雷達 QPE 演算法持續進行研究與調整，目前已加入雙偏極化雷達參數的貢獻，大幅改善雷達 QPE 的正確性。

即便以雷達 QPE 做為真值 (ground truth) 有上述不完美之處，但因傳統的自動雨量站僅分布在陸地上，而臺灣陸地僅占 WEPS 模式範圍之小部分，其校驗結果無法完整反應 WEPS 的降雨預報表現，因此本研究決定採用目前公認是臺灣近海地區最佳的 QPESUMS 雷達降雨估計做為校驗用的觀測資料。

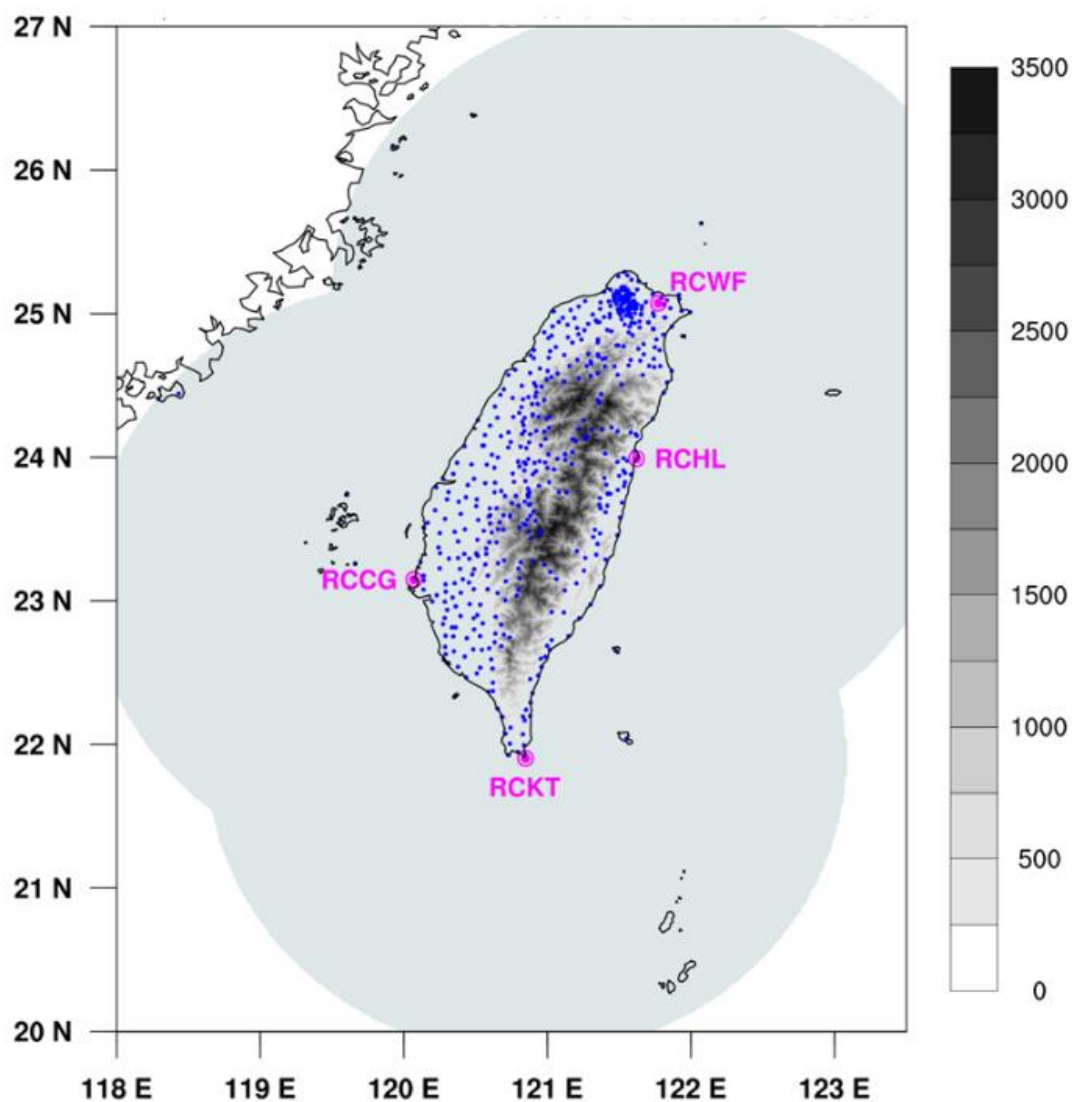


圖 2、QPESUMS 的範圍，圖中的灰藍色區域為雷達回波資料之覆蓋區域，圖中同時標示出 4 個雷達站(桃紅色圓圈)、用以校正雷達降水估計之 528 個雨量站(藍色點)以及地形高度(高度單位是公尺)。4 個雷達站分別是五分山(RCWF)、墾丁(RCKT)、花蓮 (RCHL)以及七股 (RCCG)雷達站。

(三)個案資料

本研究所使用的個案資料包括 WEPS 在 2013 ~ 2015 年共 12 個颱風個案(如表 1)的 00 - 24、24 - 48、

00 - 72 小時降雨預報。為評估 WEPS 在不同區域的預報表現及特性，我們 WEPS 模式範圍區分為全區(total area)、海洋區域(sea area)和陸地區域

(land area)；並進一步將陸地區域分為平地(plain；地形高度 < 500 m)及山區(mountain；地形高度 ≥ 500 m)。要特別說明的是：各項校驗分析均顯示海洋區域和全區兩者的差異非常小，這是因為陸地樣本僅占全部樣

本的 9%，因此全區的校驗結果近乎海洋區域的結果。為精簡篇幅，因此本論文省略海洋區域的校驗結果，僅保留 WEPS 在全區的整體預報評估結果。

表 1、2013 年至 2015 年颱風個案列表。

	Typhoon	Start - end
2013	Soulik (TY 01)	0000 UTC 11 Jul - 0000 UTC 14 Jul
	Trami (TY 02)	0000 UTC 20 Aug - 0000 UTC 22 Aug
	Kong-Rey (TY 03)	0000 UTC 27 Aug - 0000 UTC 30 Aug
	Usage (TY 04)	1800 UTC 19 Sep - 1200 UTC 22 Sep
	Fitow (TY 05)	1800 UTC 04 Oct - 0600 UTC 07 Oct
2014	Matmo (TY 06)	0000 UTC 21 Jul - 1800 UTC 23 Jul
	Fung-Wong (TY 07)	1800 UTC 18 Sep - 1200 UTC 22 Sep
2015	Linfa (TY 08)	1800 UTC 05 Jul - 0600 UTC 08 Jul
	Chan-Hom (TY 09)	1200 UTC 08 Jul - 0600 UTC 11 Jul
	Soudelor (TY 10)	1800 UTC 05 Aug - 0600 UTC 09 Aug
	Goni (TY 11)	0600 UTC 20 Aug - 1800 UTC 23 Aug
	Dujuan (TY 12)	1800 UTC 26 Sep - 1800 UTC 29 Sep

三、研究方法

(一) 系集預報評估方法

EPS 的品質好壞決定了系集機率預報的良窳。首先我們將利用系

集離散度-預報能力關係(spread-skill relationship)來評估 EPS 的品質好壞，以了解 WEPS 的系集散度能否適當的反應預報的不確定性。簡言之，亦

即評估 WEPS 在系集散度(ensemble spread) 較小的情況下，是否具有較小的預報誤差或較好的預報能力(forecast skill)。如果有這樣的關係存在，那我們可以進一步利用系集散度的大小來預測 EPS 的預報能力(forecasting forecast skill；Kalnay and Dalcher 1987)。

其次，針對 EPS 的預報表現評估，我們將採用的校驗方法包括：rank histogram (RH；Hamill 2001)、reliability diagram(又稱為 Talagrand diagram；Hsu and Murphy 1986；Hamill 1997)、ROC 面積(area under relative characteristic curve, ROC area；Mason and Graham 1999；Jolliffe and Stephenson 2003；Hamill and Juras 2006；Wilks 2011)、Brier score(BrS)、Brier skill score(BrSS；Wilks 2011)、相對均方根誤差(relative root mean square error, RRMSE)以及空間相關係數(spatial correlation coefficients)。其中，RH 可用以了解 EPS 的系集散度是否能恰當地代表實際的觀測變異。Reliability diagram 則是用以評估機率預報的系統性偏差，亦即了解天氣事件之預報機率與觀測頻率的一致性。ROC 是利用一組遞增的機率做為預警門檻(亦即，當預報機率大於或等於此門檻時，代表預報有

事件發生)，所畫出 y 軸是命中率(hit rate, HR)、x 軸是錯誤預報率(false alarm rate, FAR)的圖；ROC 曲線下方的面積(ROC 面積)可用以代表預報系統區辨事件(events)與非事件(non-events)的能力。BrS 可用以量測機率預報的誤差大小(亦即正確性)；BrSS 則是用以評估機率預報相對於氣候預報的預報技術(forecast skill)。

(二) 拔靴法與 Mann-Whitney 檢定

Chang et al. (2015)指出由於地形鎖定效應(terrain-locking effect)，在迎風斜坡上，颱風降雨的可預報度通常高於平地，且降雨機率預報的預報技術和區辨能力也相對較佳。本研究將針對此特別現象進一步探討，評估 WEPS 是否同樣在山區比在平地具有較高的可預報度以及較佳的預報能力；這部分主要是採用拔靴法(bootstrap method, Chu 2002；Chang et al. 2012)來計算校驗統計量的信心區間(confidence interval)，並採用 nonparametric Mann-Whitney test 來檢定此現象是否具有統計上的顯著性。

1. 信心區間與拔靴法

信心區間是利用所取得的樣本，對母體(population)的某個測試統計量(例如平均數、中位數或標準差等)進行區間估計，以瞭解此測試統計量其樣本量測值的可信程度，並用以

描述樣本對母體統計量估計的不確定性。如何建立信心區間呢?首先需選定一個信心程度(confidence level)，例如 95%，此信心程度代表信心區間涵蓋母體參數的機率。簡言之，信心區間考慮了樣本變異以及樣本數目有限所造成的統計不確定性，因而可對預報表現提供更具參考價值的資訊。

在本研究中，我們是採用拔靴法(bootstrap method；Wilks 2011)，透過重新取樣(resampling)來建立信心區間。拔靴法的概念在於每個樣本是獨立的，因此每個樣本被抽到的機率是相等的。本研究共有 135 個個案(樣本數 $N=135$)，從總樣本中抽取 135 次以重新得到 135 個樣本(同一樣本可以重複被抽取)，即可產生一個拔靴樣本(在本研究中乃是從重新取樣得到的 135 個樣本中，得到一個相關係數中位數)；透過反覆的重新取樣過程，我們可以得到大量的拔靴樣本(本研究的拔靴樣本數是 10,000)，以建立母體統計參數(本研究是中位數)的抽樣本分佈(sampling distribution)。接著利用分位法(percentile method)找出信心區間，例如本研究是找出拔靴樣本分布的第 5 和 95 百分位，以建立相關係數中位數的 90%信心區間。

2. Mann-Whitney 檢定

Mann-Whitney test (Mann 1945；Larson 1982) 屬於無母數(nonparametric)統計檢定，適用於母體分佈未知、樣本數少以及母體不是常態分布，也不易轉換為常態的狀況。Mann-Whitney test 又稱為 two-sample rank test，雖然 Mann-Whitney test 並未要求所測試的資料必須是常態分布，但必須滿足一個關鍵性的假設：樣本必須是獨立的(independent)，亦即樣本彼此間的相關性非常小。因此檢定前，必須先檢查樣本的測試統計量(本研究是以 median correlation 做為測試統計量)在不同時滯(lag)的自相關函數(autocorrelation function)，以確定樣本是彼此獨立的。之後，再選定顯著水準(level of significance, α)，利用 Mann-Whitney test 來進行下列的假設測試：

虛無假設 H_0 ：

$$\text{median correlation}_1 = \text{median correlation}_2$$

對立假設 H_A ：

$$\text{median correlation}_1 \neq \text{median correlation}_2$$

其中，下標 1 和 2 代表待檢定的兩組樣本，在本研究中分別是山區和平地兩組樣本。Mann-Whitney test 是透過下列的 Z_W 公式將資料轉換成常態近似來決定 p 值(p-value)。

$$(1) \quad Z_w = \frac{\left| W - \frac{n(m+n+1)}{2} \right| - 0.5}{\sqrt{\frac{mn}{12} \left[(m+n+1) - \frac{\sum_{i=1}^K (t_i^3 - t_i)}{(m+n)(m+n-1)} \right]}}$$

其中 n 和 m 分別是樣本 1(山區)和樣本 2(平地)的個數， K 是重複值(tied value)的組數， t_i 是第 i 組重複值的個數。透過以下 4 個步驟，可以求出 W 值：(1)將兩組樣本集中在一起，(2)將所有的 correlation 值由小到大排序，(3)找出 correlation 值相同的樣本，這些樣本的序位(rank)是採用平均序位，比方說，序位 11 到序位 13 的 correlation 值均相同，則這幾個 correlation 值的序位將重新被給定為 12 [= (11+12+13)/3]，(4)計算樣本 1 的序位總合，此即為 W 值。

利用公式(1)求出 Z_w 值之後，再針對常態近似下的雙邊測試(two-sided test)求出 p 值； p 值代表山區和平地的預報能力(本研究是以 median correlation 做為預報能力指標)沒有差異的機率。倘若 p 值 $\leq \alpha$ ，則拒絕虛無假設，亦即山區和平地的預報能力在統計上具有顯著差異；倘若 p 值 $> \alpha$ ，則接受虛無假設，代表山區和平地的預報能力在統計上沒有差異。

四、結果與討論

(一) 系集散度

在本研究中，我們以系集平均之方均根誤差(root mean square errors, RMSE)做為預報誤差之量測，並以系集散度(ensemble spread, SPRD)做為散度之量測(Whitaker and Loughe 1998)。圖 3 顯示系集降雨預報(0-24、24-48 以及 0-72 小時預報)之 SPRD 與 RMSE 具有高度線性相關，相關係數都高於 0.84，且決定係數(coefficient of determination, R^2)均高於 0.7；顯示系集散度可以良好地反應預報的不確定性和預報誤差。但要注意的是，當系集平均的 RMSE 較小(代表天氣系統的可預報度較大)時，WEPS 的散度稍微偏小(underdispersive；落在散布圖對角線上方的點)；而當系集平均的 RMSE 較大(代表天氣系統的可預報度較小)時，WEPS 的散度明顯偏大(overdispersive；落在散布圖對角線下方的點)，且散度偏大的情狀會隨著天氣個案的可預報度減小而愈加明顯。要特別說明的是：在本研究中，陸地樣本只占全部樣本的 9%，因此整個模式範圍的結果幾乎等同於海洋區域的結果。

針對颱風系集降雨預報，颱風

路徑的正確與否對於預報結果的良窳扮演決定性的角色。在部分不確定較大的個案(例如颱風移入中國大陸或是颱風減弱消散)中，可以發現 20 個系集成員的颱風路徑相當分歧，繼而導致降雨與風速預報的系集散度過大。但在絕大部分的個案中，系集散度都可以良好地反應預報的不確定性和預報誤差，換句話說，WEPS 無論是在陸地或海洋區域，均可產生具有良好品質的系集降雨預

報，特別是 0-24 小時以及 24-48 小時的系集降雨預報。因此，接下來我們將透過 WEPS 發展系集降雨和風速機率預報。圖 4 是針對 2013 年侵臺的天兔(Usagi)颱風，所產生在不同降雨門檻的系集降雨機率預報(左圖)及相對應的觀測降雨機率(右圖)，觀測降雨取自雷達降雨估計。可以發現，WEPS 在天兔颱風個案中，預報和觀測的降雨位置有相當不錯的對應。

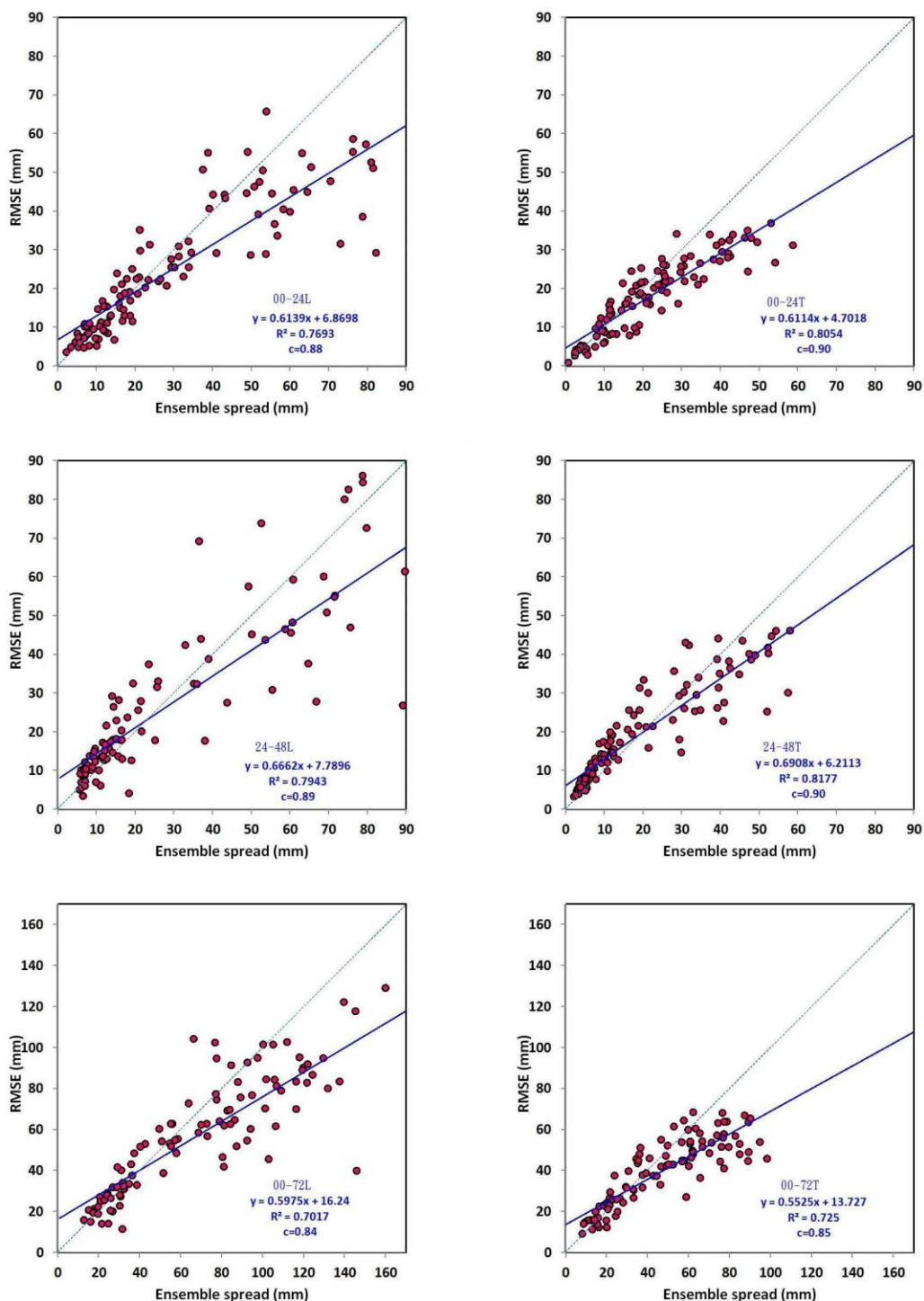


圖 3、系集散度(ensemble spread)與系集平均之方均根誤差(RMSE)的散佈圖，校驗之颱風個案請參見表 1。散布圖上的每一個點來自於一個 0 - 24 小時(第一列)、24 - 48 小時(第二列)或 0 - 72 小時(第三列)系集降雨預報的結果(亦即，ensemble spread 和 RMSE 是模式校驗範圍之平均)。左欄是陸地區域，而右欄是整個模式範圍的結果。圖上同時標示出線性迴歸線(深藍色實線)、相關係數(C)以及決定性數(R^2)。

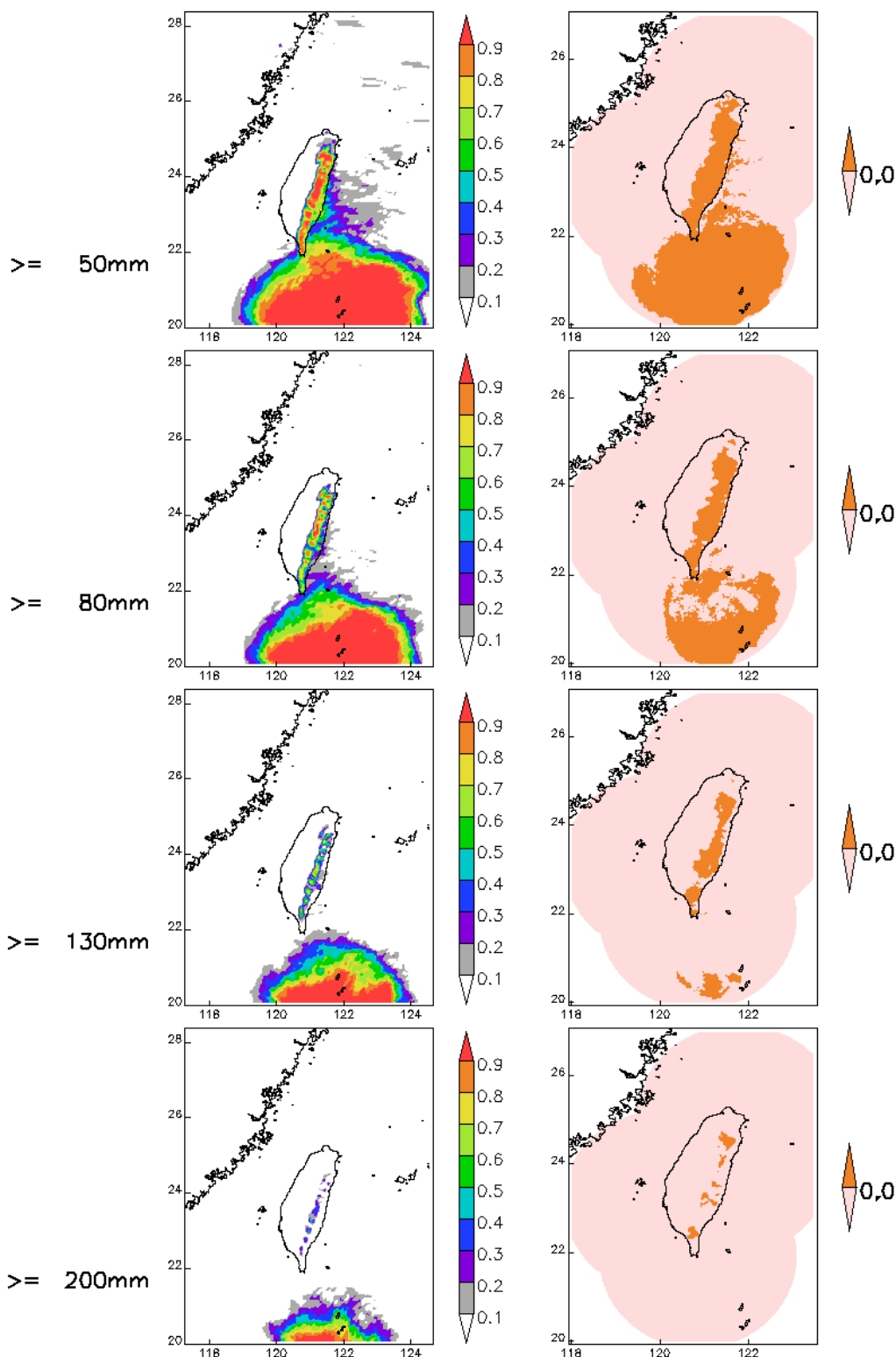


圖 4、WEPS 0 - 24 小時系集降雨機率預報(左圖)以及所對應的觀測降雨機率(右圖)在不同門檻 [50 、 80 、 130 和 $200 \text{ mm (24 h)}^{-1}$] 的分布，個案結束時間是 2013 年 9 月 21 日 0600 UTC。右圖中的深橘色區域代表雷達降雨估計達到降雨門檻(亦即，觀測降雨機率為 1)，而粉橘色區域是中央氣象局 4 顆作業用雷達(包括五分山、花蓮、七股和墾丁雷達)的覆蓋區域。

(二) 預報偏差

1. rank histogram

Rank histogram (RH)可用以了解EPS的預報偏差或是評估系集散度是否能恰當地代表實際的觀測變異。圖 5(a)(b)是利用本研究 12 個颱風個案(表 1)之 0-24 小時系集降雨預報進行統計所得到在陸地區域和全區的 RH(24-48 和 0-72 小時的結果類似，因此未附圖)。RH 的做法是針對每一筆統計樣本(包含一個觀測降雨和所有系集成員的預報降雨)，將觀測降雨和所有系集成員(20 個成員)的預報降雨進行排序，由最小(rank=1)排到最大(rank=21)，然後統計觀測降雨落在每一順位(rank)的發生頻率即可得到 RH。可以發現：陸地區域和全區的 RH 均呈現 L 型，有很高比率的統計樣本(海洋區域明顯高於陸地區域)其觀測序位最小，亦即這些樣本其所有系集成員的預報降雨均大於觀測降雨，顯示 WEPS 有明顯的濕偏差，且海洋區域濕偏差的情況較陸地區域顯著。海洋區域濕偏差特別顯著的原因在於：本研究採用雷達 QPE 做為觀測，但此資料可能在海上有明顯低估的狀況，造成校驗結果中 WEPS 呈現假性的過度濕偏差。

除了雷達 QPE 在海上有低估的因素外，WEPS 整體呈現明顯濕偏差的

主要原因是模式在小雨($< 1 \text{ mm}$)時會有過度預報的系統性偏差，例如觀測無雨時，模式可能會有 0.5 mm 的預報降雨，這和模式的微物理過程有關，因此氣象局預報員以 WEPS 做為有雨/無雨的預報參考時，都會進行主觀的偏差修正。圖 5(a)(b)的 RH 分析顯示 WEPS 有明顯的濕偏差，事實上觀測序位最小的樣本(陸地區域~28%；全區~39%)中，很多是來自「觀測無雨」的樣本，但這些樣本的預報降雨卻介於 0-1 mm 之間。因此當我們拿掉觀測無雨的樣本後再進行 RH 分析[圖 5(c)(d)]，可以發現 WEPS 整體呈濕偏差的現象大幅改善了，陸地和海洋區域中觀測序位最小的樣本，只占有「觀測有雨」樣本的 10.9%和 10.4%。

2. reliability diagram

Reliability diagram 主要用以評估系集機率預報的系統性偏差，亦即了解預報機率與觀測頻率的一致性。圖 6 是 0-24 小時系集降雨機率預報在不同降雨門檻下的 reliability diagram，圖中的曲線是信賴曲線(reliability curve)，而對角線代表完美預報。信賴曲線和對角線愈接近表示機率預報的偏差愈小，可信度愈高。可以發現：當預報機率較小(低於 30%)時，有輕微的低報；而當預報機率較大

(約高於 50%)時，會逐漸偏向高報(濕偏差)。整體而言，可信度隨著降雨門檻的提高以及預報時間的增加而降低(24-48 小時以及 48-72 小時的圖未附)；此外，陸地區域的可信度明顯優於海洋區域，這與 rank histogram

的分析結果是一致的，海洋地區的濕偏差比陸地區域顯著。如前所述，這很可能是因為做為觀測的雷達降雨估計在海洋區域低估所導致。

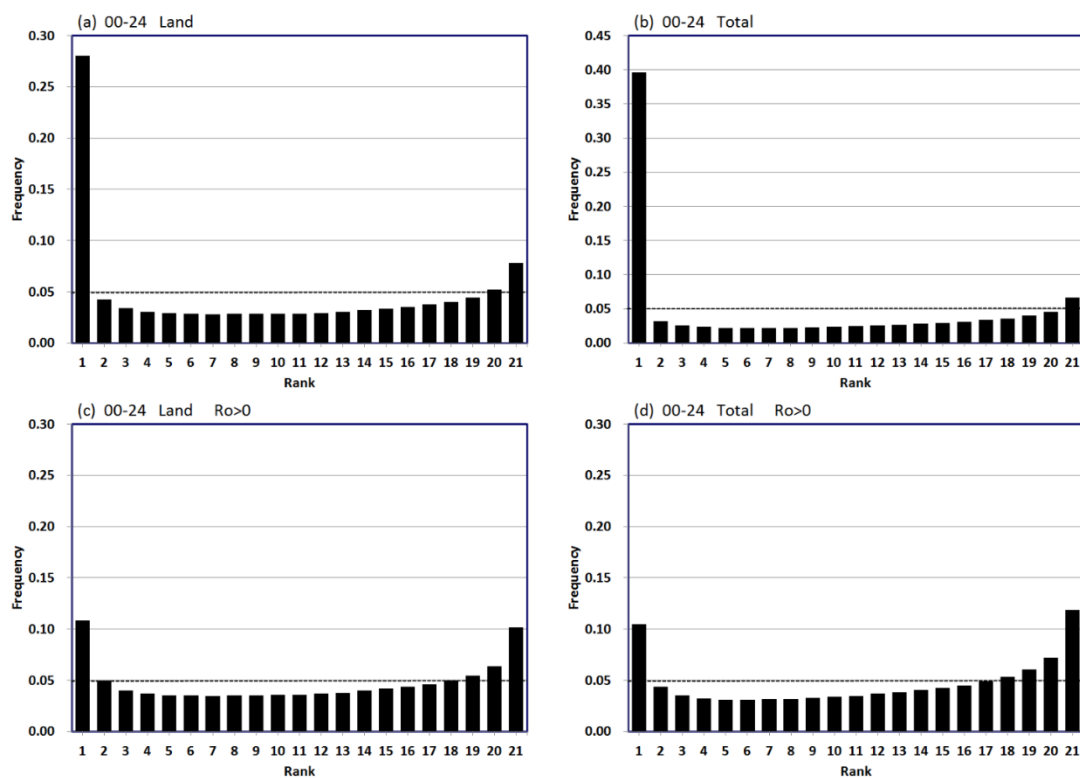


圖 5、WEPS 0 - 24 小時(c)(d) 系集降雨預報在(a)陸地區域和(b)整個模式範圍的 rank histogram；(c)(d)同(a)(b)，但只納入觀測有雨的樣本。圖中的水平虛線代表觀測降雨之順位(rank)均勻分布在所有順位(1~21)之狀況下的頻率。

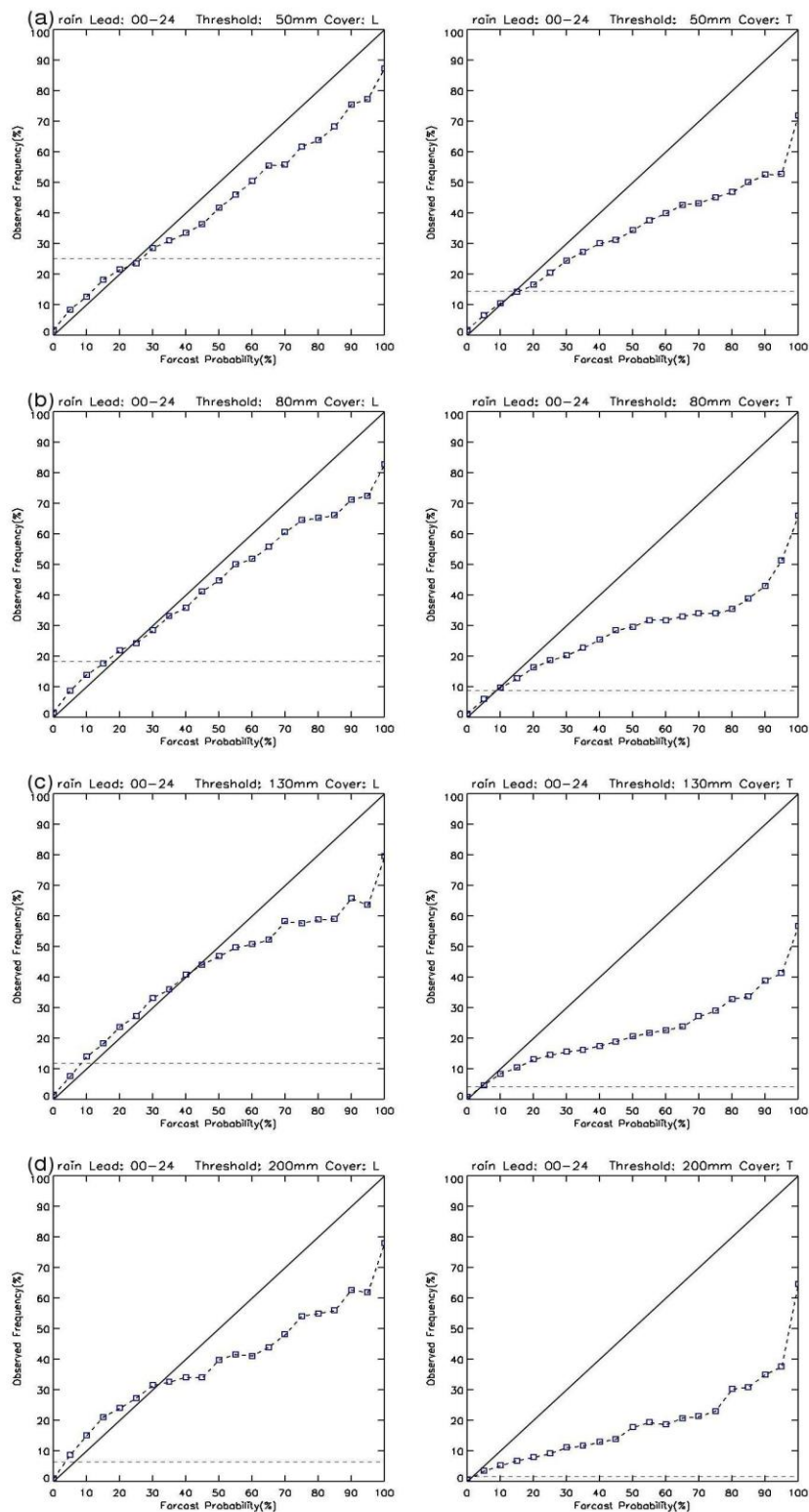


圖 6、WEPS 0-24 小時 PQPFs 在不同降雨門檻[(a) 50、(b) 80、(c) 130 以及 (d) 200 mm (24 h)⁻¹] 下的 reliability diagrams，左欄是陸地區域的結果，而右欄是整個模式範圍的結果。Reliability diagrams 中的水平虛線代表機率預報相對於氣候不具有區辨能力(no resolution)，亦即觀測頻率等於氣候頻率之狀況。

(三) 區辨能力

ROC 是利用一組遞增的機率做為預警門檻(亦即, 當預報機率大於或等於此門檻時, 代表預報有事件發生), 所畫出 y 軸是命中率 (hit rate, HR)、x 軸是錯誤預報率 (false alarm rate, FAR)的圖(如圖 7)。應用上主要以 ROC 曲線下方的面積(ROC 面積)大小來評估預報系統區辨事件(events)與非事件(non-events)的能力, 當 ROC 面積 >0.7 時, 代表模式具有好的區辨能力(Buizza et al. 1999)。亦即, 針對觀測到的事件, 模式有很高的頻率預報有事件發生, 且針對觀測到的非事件, 模式有很高的頻率預報無事件發生。舉例來說, 當我們根據觀測結果將所有樣本區分為大雨(24 小時累積雨量 ≥ 80 mm)和非大雨事件時, 模式對於大雨事件有很高的頻率預報有大雨發生, 且對於非大雨事件有很高的頻率預報無大雨發生; 換句話說, 針對大雨事件, 模式在各預警門檻的命中率相對很高, 且錯誤預報率相對很低。

分析結果顯示: WEPS 系集降雨機率預報在不同區域(海洋區域、山區和平地)、不同降雨預報時段(0-24、24-48 和 48-72 小時)的 ROC 面積均 >0.8 (圖 8), 亦即 WEPS 具有良好的

區辨能力。此外, 陸地區域的區辨能力優於海洋區域(ROC 面積較大), 且山區(地形高度 ≥ 500 m)的區辨能力優於平地(地形高度 <500 m); 隨著預報時間增加以及降雨門檻提高, 區辨能力會逐漸降低。

此處再次釐清可信度與區辨能力的差異, 可信度要回答的問題是: 當模式預報有事件發生時, 實際發生的狀況如何?而區辨能力要回答的問題是: 假定觀測到有事件發生, 那麼其所對應的預報結果如何呢?可信度與區辨能力是彼此獨立(independent)的兩種特性(attribute)(Cartwright 1956), 亦即, 可信度的高低與區辨能力的好壞並不互相影響, 因此一個有預報偏差的預報系統, 仍然可能有好的區辨能力(Chang et al. 2012; Chang et al. 2015)。機率預報中的區辨能力和可信度的關係類似決定性預報中的空間相關係數與預報偏差的關係; 舉個簡單的例子來說, 某模式降雨預報有完美的相關係數($=1$; 代表模式與觀測的降雨型態相同), 但卻可能有嚴重的預報偏差(例如模式範圍內每個格點的預報降雨量均是觀測降雨量的 3 倍)。因此, ROC 面積只能做為預報系統潛在可用性(potential usefulness)的量測, 實際上預報系統

的可用性會因為預報偏差而受影響；必須透過校正程序修正預報偏差後，才能達到潛在可用性。預報校驗結果顯示 WEPS 系集降雨機率預報具

有良好的區辨能力，但具有明顯的預報偏差，這意味著透過後端校正程序，可以大大提升系集降雨機率預報的能力。

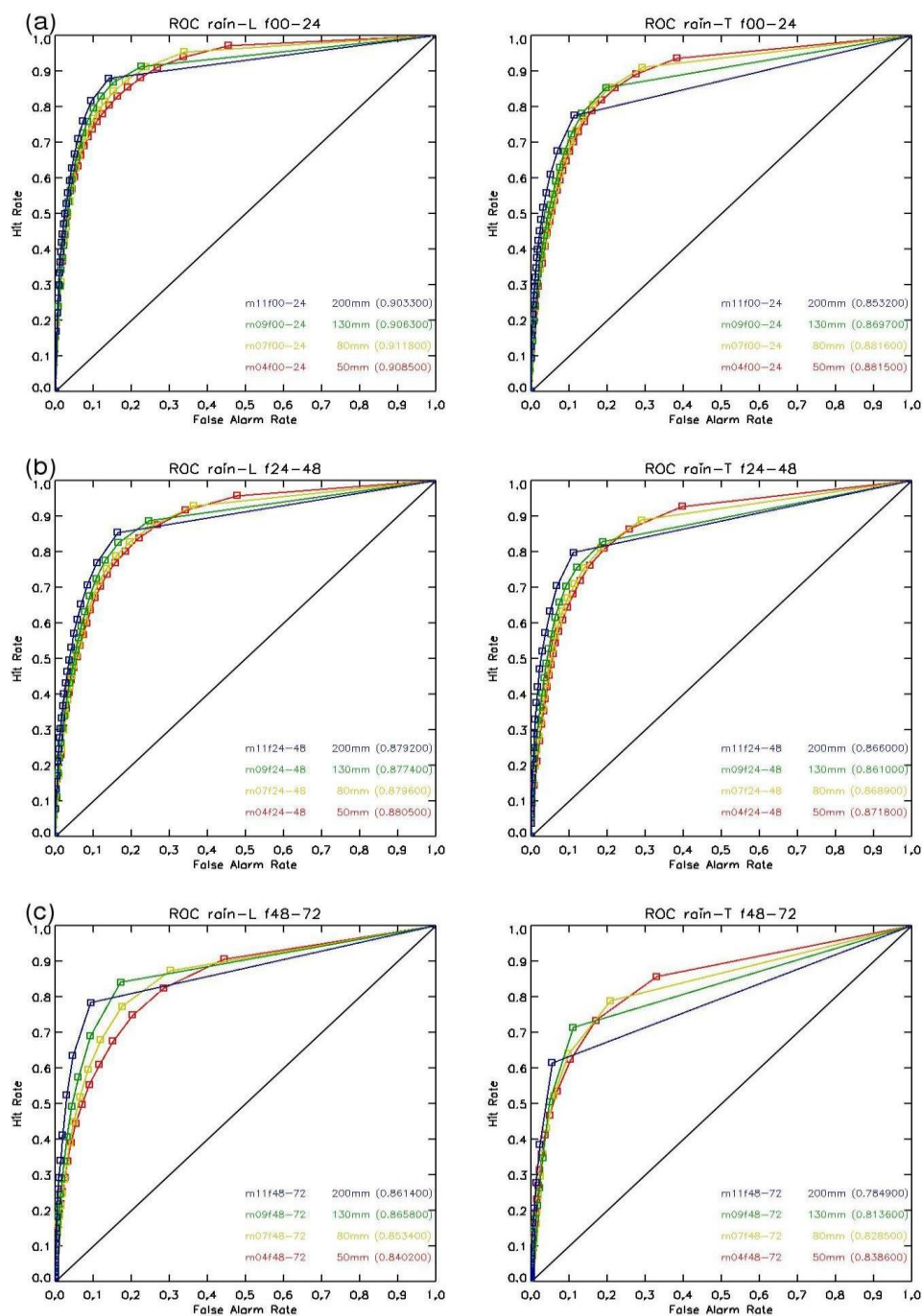


圖 7、WEPS (a) 0 - 24 小時、(b) 24 - 48 小時和(c) 48 - 72 小時在不同降水門檻[50、80、130 和 200 mm (24 h)⁻¹]下的 ROC 曲線。

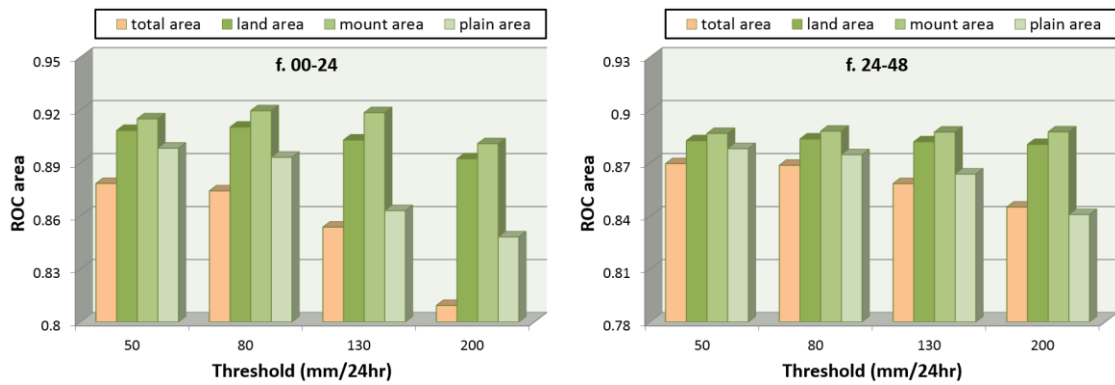


圖 8、0 - 24 小時(左圖)以及 24 - 48 小時(右圖)系集降雨機率預報在不同門檻 [50、80、130 和 200 mm (24 h)⁻¹]與不同區域(模式全區、陸地區域、山區和平地)的 ROC 面積。

(四) 預報技術

我們進一步計算 Brier skill score ($BrSS$)，來評估系集機率預報相對於氣候預報的誤差改善程度；亦即，相較於使用氣候資訊進行降雨預報，系集機率預報否具有預報技術。

$$BrSS = \frac{BrS - BrS_{ref}}{BrS_{perf} - BrS_{ref}} = 1 - \frac{BrS}{BrS_{ref}} \quad (2)$$

其中 Brier score (BrS) 是預報系統的機率預報誤差大小，其定義如下

$$BrS = \frac{1}{N} \sum_{j=1}^N (P_j - O_j)^2 \quad (3)$$

N 是所有的校驗樣本數， P_j 是預報機率， O_j 是觀測頻率。而在方程式(2)中， BrS_{ref} 是 BrS 的氣候參考值，亦即以氣候降雨頻率做為預報降雨機率所求出的 BrS 。氣候降雨頻率通常是採用長期氣候值或樣本氣候值，本研究是採用樣本氣候值； BrS_{perf} 是

完美預報系統的 BrS ，因此 $BrS_{perf}=0$ 。 $BrSS$ 的值域從 $-\infty$ 到 1； $BrSS=0$ 代表機率預報相對於氣候預報是沒有預報能力的， $BrSS=1$ 代表完美預報， $BrSS > 0$ 代表機率預報相對於氣候預報是有預報能力的。

圖 9 是 WEPS 0-24 小時系集降雨機率預報於全區、陸地、山區和平地不同門檻下的 $BrSS$ ，可以發現：相對於氣候預報，陸地區域在門檻高達 350 mm (24 h)⁻¹ 的降雨事件中依然具有預報技術 ($BrSS > 0$)，但海洋區域(或全區)對於門檻高達 130 mm (24 h)⁻¹ 的降雨事件就不再具有預報技術。

(五) 山區和平地的雨量預報表現

比較

地形效應深深影響臺灣地區的

颱風降雨，高達 3800 m 的中央山脈對颱風降雨的空間分佈扮演著重要的角色(Yeh and Elsberry 1993)，亦即颱風降雨的空間分布主要決定於颱風中心和中央山脈的相對位置，此即所謂的地形鎖定效應 (terrain-locking effect；Su et al. 2012)。由於地形鎖定效應，颱風引發之大降雨經常發生在迎風面的斜坡上(地形高度 ≥ 500 m)；因此，在迎風斜坡上，颱風降雨的可預報度通常高於平地，且預報系統的預報技術和區辨能力也相對較佳(Chang et al. 2015)。本研究針對此現象，評估 WEPS 是否同樣在山區比在平地具有較佳的預報能力？倘若預報能力具有地域上的差異，我們將進一步採用 Mann-Whitney test 來檢定山區和平地的預報能力差異是否具有統計上的顯著性。

1. 可信度、區辨能力與預報技術

比較山區和平地的可信度曲線(圖 10)，可以發現：WEPS 在山區呈現濕偏差，在平地，當預報機率低於 50%時，WEPS 呈現些微的乾偏差；而當預報機率高於 50%時則是呈現些微的濕偏差。整體而言，WEPS 系集降雨機率預報在平地的可信度優

於山區。WEPS 在山區的濕偏差高於平地，有可能與採用雷達 QPE 做為觀測真值有關。由於陸地區域未經校正的雷達 QPE 主要呈現低估現象，因此在雨量計稀少的山區，雷達 QPE 很可能會低估。但山區較高的濕偏差是否主要來自於雷達 QPE 的低估，有待進一步的探討。此外，WEPS 區辨能力分析(圖 8)顯示：山區的區辨能力優於平地，特別是對於較大降雨事件。而預報技術分析(圖 9)顯示：在各降雨門檻 $[30\sim 350 \text{ mm (24 h)}^{-1}]$ ，山區的預報技術均優於平地，亦即有較高的 BrSS。由於預報技術主要來自可信度與區辨能力的共同貢獻；換句話說，山區在區辨能力上對於預報技術的正貢獻，大於其在可信度上對於預報技術的負貢獻。對於較大降水事件，WEPS 在山區之區辨能力優於平地的結果同 Chang et al. (2015)；這是因為中央山脈的地形鎖定效應，造成颱風引發之大降雨經常發生在迎風面的斜坡上(地形高度 ≥ 500 m)；亦即，在迎風斜坡上，颱風降雨的可預報度通常高於平地，因此模式降雨預報的區辨能力相對較佳。

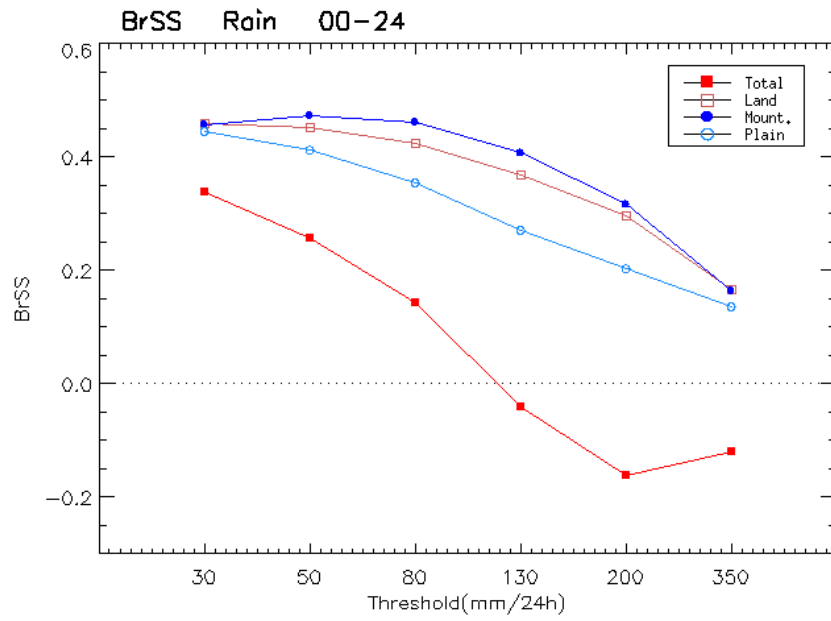


圖 9、WEPS 系集降雨機率預報在不同區域(模式全區、陸地、山區和平地)以及不同門檻[35、50、80、130、200 以及 350 mm (24 h)⁻¹]下的 BrSS。

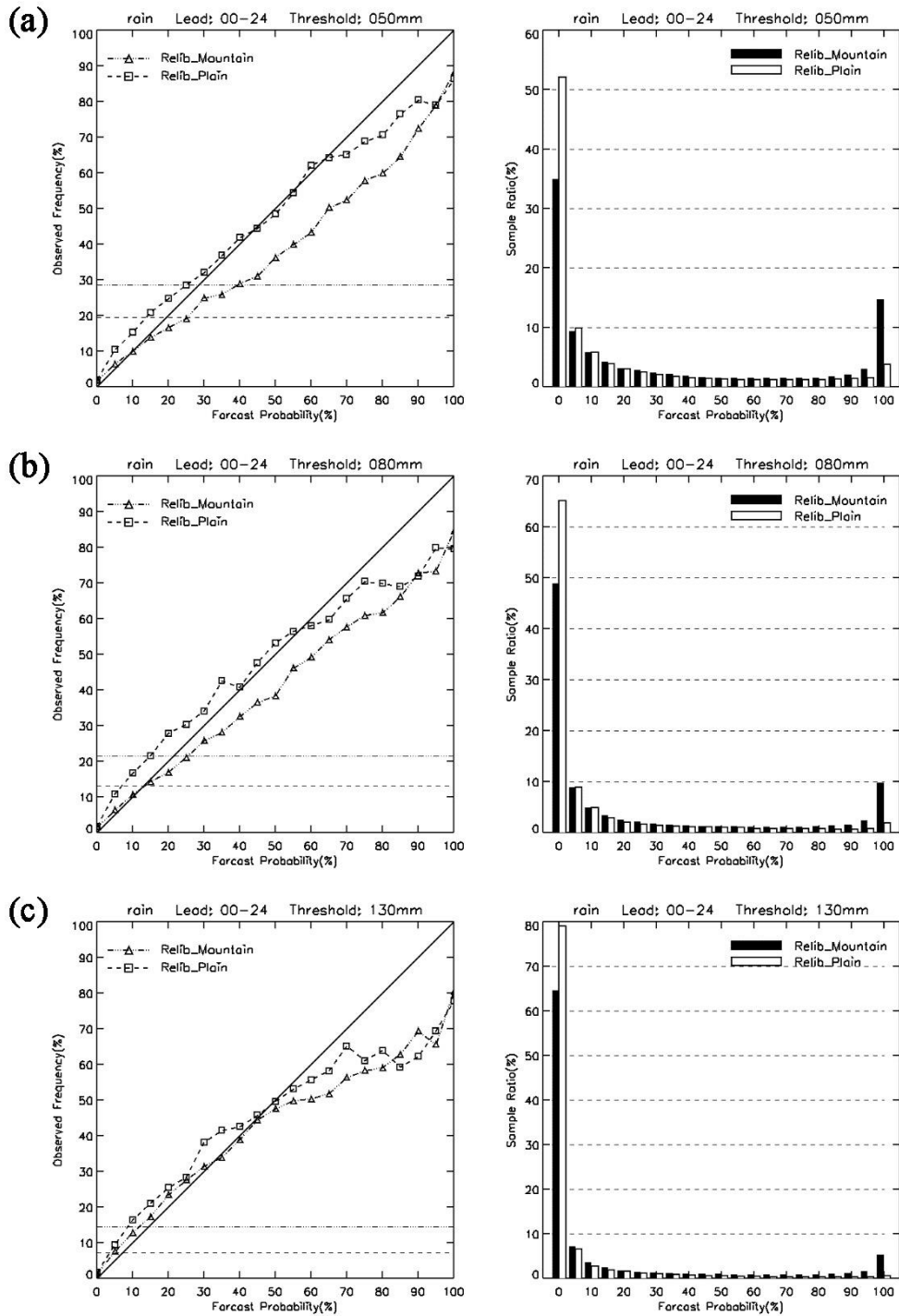


圖 10、WEPS 0-24 小時 PQPFs 在不同降雨門檻[(a) 50、(b) 80 以及 (c) 130 mm (24 h)⁻¹] 下的 reliability diagrams (左欄) 以及相對應的 histograms (右欄)，三角形標記與黑色長條圖是山區的结果，而方形標記與空白長條圖是平地的结果。Reliability diagrams 中的水平虛線代表氣候發生頻率，針對同一門檻，山區的氣候發生頻率高於平地。

2. 空間相關係數與相對均方根誤差 (RRMSE)

上述之可信度、區辨能力與預報技術(BrSS)分析都是針對我們所關心的事件來進行分析，也就是針對不同降雨門檻來評估模式的條件性預報表現，而非整體的預報表現。接下來我們將透過空間相關係數與預報誤差分析來評估模式的整體預報表現。

針對 WEPS 0-24 小時降雨預報，20 個系集成員與系集平均其預報與觀測之相關係數及 RRMSE 的盒圖(boxplot)分析(圖 11)顯示：就中間 50%的資料分布而言，各成員在山區的相關係數相對高於平地，但山區 RRMSE 的離散度相對較大，且山區 50-75 百分位的 RRMSE 明顯較大。換句話說，WEPS 在山區的降雨預報有較佳的降雨型態(rainfall pattern)，但因山區的預報偏差比較大，其 RRMSE 未必比較低。此外，系集平均的相關係數較個別成員高，且 RRMSE 較個別成員低；亦即系集平均的預報表現優於個別的成員。這是因為系集平均可以濾掉非線性誤差，而得到比個別成員更佳的預報表現。

盒圖分析同時顯示相關係數並非呈現常態分布(因降雨是高度非線

性的事件)，因此本研究採用中位數而非平均數來代表其中間趨勢(central tendency)，並以相關係數的中位數做為測試的統計量(test statistic)，利用拔靴法建立 95%的信心區間。分析山區和平地之相關係數差值的中位數及其 95%信心區間(圖 12)發現：不管是個別成員或系集平均，山區和平地之相關係數差值的中位數均 > 0 ，且其 95%信賴區間的值也都 > 0 ；換句話說，山區所預報的降雨型態比平地更為正確。此外，個別成員的 95%信賴區間較系集平均大；這表示相較於系集平均，個別成員在山區和平地的預報表現有較大的變異。

為進一步了解山區和平地的預報表現差異是否具有統計上的顯著性，我們以相關係數做為預報能力指標來進行統計推論。這是因為相關係數反映降雨型態的正確與否，但 RRMSE 除了反映降雨型態的正確與否外，還夾雜了預報偏差的影響。針對預報偏差，我們可以輕易地透過後端校正加以修正，但降雨型態的正確預報與否主要決定於模式的區辨能力，這部分必須藉由改進模式的物理才能獲得改善。

我們以相關係數的中位數做為測試統計量，並選定顯著水準[level

of significance, α ; 本研究採用 90 % 信心區間, 亦即 $(1-\alpha)\%=90\%$, 因此 $\alpha=0.1$], 利用 Mann-Whitney test 來進行下列的雙邊檢定(two-sided test)。

虛無假設 H_0 : median correlation₁ = median correlation₂

對立假設 H_A : median correlation₁ \neq median correlation₂

其中, 下標 1 和 2 分別代表山區和平地。測試結果顯示: 20 個系集成員中, 有 19 個成員的 p 值 ≤ 0.1 , 代表我們可以拒絕虛無假設, 而接受對立假設; 亦即有 95%(=19/20)的系集成員其山區的預報能力優於平地具有統計上的顯著性。

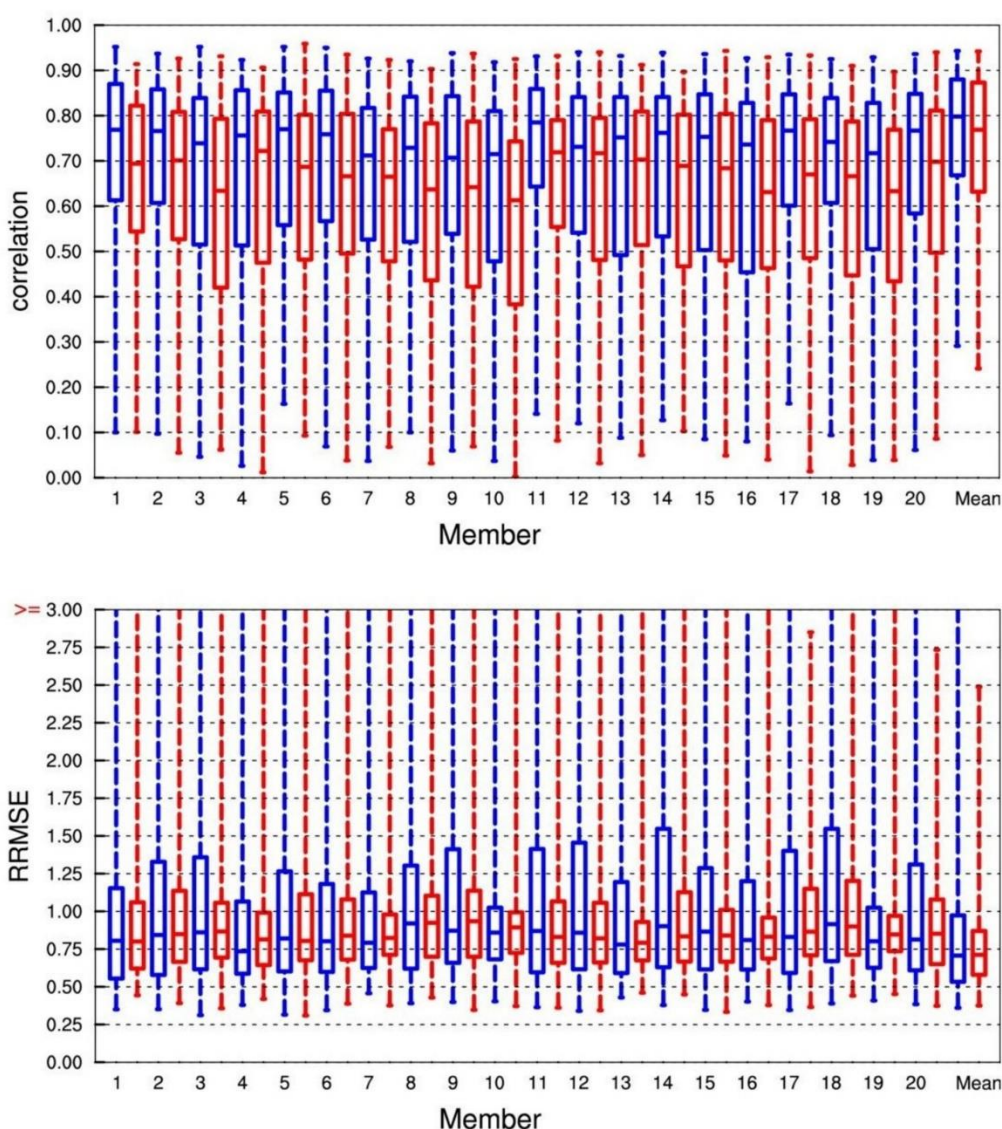


圖 11、WEPS 20 個系集成員(1~20)與系集平均(ENS)在山區(藍線)和平地(綠線) 0 - 24 小時降雨預報的相關係數(correlation; 上圖)與相對的均方根誤差 (RRMSE; 下圖)。

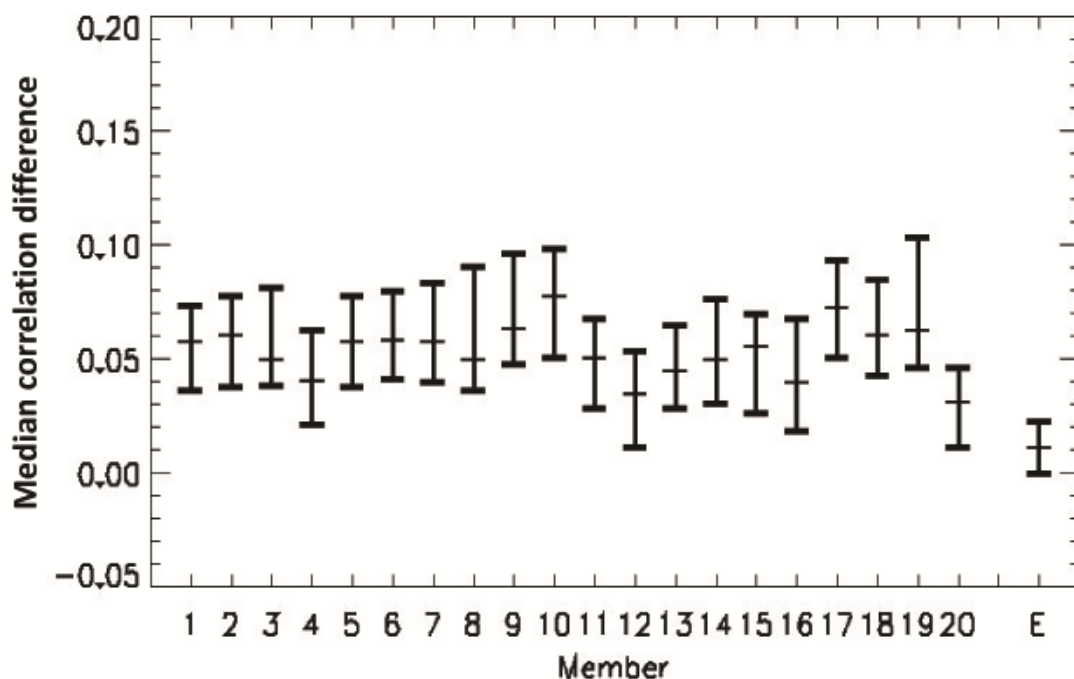


圖 12、WEPS 20 成員(1~20)和系集平均(E)在山區和平地之相關係數差值的中位數(median correlation difference)，以及利用 bootstrap 方法(重新取樣 10,000 次)得到的 95%信心區間。

五、 結論與展望

本研究主要是針對臺灣地區的颱風，利用中央氣象局發展的WEPS，進行定量的降雨機率預報；最終目的在於提供更具參考價值的系集颱風降雨預報產品，並讓使用者藉由系集機率預報做出最佳化的決策，以得到最大的經濟價值。

本論文(第一部份)的研究重點在於評估WEPS的品質良窳，並利用WEPS產生颱風降雨機率預報，詳細校驗其預報表現。研究結果顯示：針對颱風降雨預報，WEPS的系集散

度與系集平均之RMSE具有高度線性相關，相關係數高於0.8，且決定係數(R^2)均高於0.7；代表WEPS的系集散度可以良好地反應預報的不確定性和預報誤差。整體而言，颱風降雨機率預報有高報情形，可信度隨著降雨門檻提高而降低，但具有相當良好的區辨能力，這意味著我們可以透過後端校正程序，大大提升WEPS系集降雨機率預報的能力。因此，我們針對系統性預報偏差，分別採用多重線性迴歸方法和非線性的類神經網路技術進行校正，並透

過敏感度實驗來探討影響校正結果的重要因子，此部分成果將發表於第二篇論文。

不同區域的校驗結果顯示：針對系集降雨機率預報，平地的可信度優於山區，但山區的區辨能力優於平地；這反映了WEPS降雨預報在山區有較佳的降雨型態，但有明顯的濕偏差。整體而言，山區的預報能力優於平地，這是因為中央山脈的地形鎖定效應導致颱風引發之大降雨經常發生在迎風面的斜坡上(地形高度 ≥ 500 m)；因此，在迎風斜坡上，颱風降雨的可預報度通常高於平地，因而WEPS在山區的預報技術和區辨能力相對較佳。我們進一步透過Mann-Whitney測試來檢定山區的預報能力優於平地是否具有統計上的顯著性。由於相關係數主要反映所預報之降雨型態是否正確，因此我們以相關係數做為預報能力指標進行假設測試，並選定顯著水準 $\alpha = 0.1$ 進行雙邊測試。檢定結果顯示：95%的系集成員，其山區的預報能力優於平地在統計上具有顯著性。

值得注意的是，降雨預報在陸地區域的可信度與區辨能力均優於海洋區域，這與做為真實場的觀測資料特性有關。本研究是以雷達

QPE做為觀測資料，陸地區域的雷達QPE有經過雨量計的校正，但海洋地區則無法校正；此外，距離雷達站較遠的海洋地區，雷達波束離地面較高，因此雷達QPE低估的可能性相當高，這是造成WEPS在海洋地區有顯著濕偏差(可信度不佳)的主要原因。至於降雨預報在陸地區域的區辨能力優於海洋區域主要還是因為地形鎖定效應的貢獻。

本研究的颱風降雨機率預報校驗涵蓋不同區域(海洋與陸地以及山區與平地)的各種預報特性(預報偏差、區辨能力、預報誤差、預報技術等)評估，目的是為了建立完整的預報評估。為使預報評估結果能夠提供預報員或使用者更有用的預報指引，未來還可進一步統計分析WEPS降雨機率預報在不同綜觀環境場、颱風路徑、移速或強度下的預報特性，比方說，不同的颱風路徑會造成預報偏差在地域上有何不同的分布？不同的颱風移速或強度是否會對預報的可信度或區能力造成明顯的影響？除了提供更有用的預報指引外，WEPS在不同分類條件下的預報偏差資訊，也將幫助我們根據不同的預報偏差分布建立不同的校正關係式，之後即可根據颱風的不同分類採用不同的校正關係式，以得到較佳的

校正結果。

致謝

作者深深感謝中央氣象局資訊中心提供WEPS降雨預報資料，以及衛星中心提供QPESUMS雷達降雨估計資料和電腦計算資源。感謝科技部自然司之研究經費補助，本研究

是在專題研究計畫補助下完成(MOST-104-2625-M-052-004、MOST-105-2625-M-052-001 和 MOST-106-2625-M-052-001)；楊舒芝由MOST 106-2625-M-008-015所支持。同時感謝兩位審查者提供寶貴的建議，使論文內容更為嚴謹且豐富。

參考文獻

- 陳怡儒、洪景山，2017：應用 decaying average方法進行地面溫度預報偏差修正。大氣科學，**45**，25 – 42。
- Atger, F., 2003: Spatial and interannual variability of the reliability of ensemble-based probabilistic forecasts: Consequences for calibration. *Mon. Wea. Rev.*, **131**, 1509 – 1523.
- Buizza, R., M. Miller, and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. *Q. J. R. Meteorol. Soc.*, **125**, 2887 – 2908.
- _____, P. L. Houtekamer, Z. Toth, G. Pellerin, M. Wei, and Y. Zhu, 2005: A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. *Mon. Wea. Rev.*, **133**, 1076 – 1097.
- _____, M., Leutbecher, and L., Isaksen, 2008: Potential use of an ensemble of analyses in the ECMWF Ensemble Prediction System. *Q. J. R. Meteorol. Soc.*, **134**, 2051 – 2066.
- Chang, H. L., H. Yuan, P. L. Lin, 2012: Short-Range (0-12h) PQPFs from Time-Lagged Multimodel Ensembles Using LAPS. *Mon. Wea. Rev.*, **140**, 1496 – 1516.
- _____, 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.

- ___, B.G. Brown, P. S. Chu, Y. C. Liou, and W. H. Wang, 2017: Nowcast Guidance of Afternoon Convection Initiation for Taiwan. *Wea. Forecasting*, **32**, 1801 – 1817.
- Cartwright, D.S., 1956: Non-parametric Estimate of Multi-judge Reliability. *Psychometrika*, **21**: 17. <https://doi.org/10.1007/BF022890842>
- Chu, P.-S., 2002: Large-scale circulation features associated with decadal variations of tropical cyclone activity over the Central North Pacific. *J. Climate*, **15**, 2678 – 2689.
- ___, Zhao X., C.-T. Lee, and M.-M. Lu, 2007: Climate prediction of tropical cyclone activity in the vicinity of Taiwan using the multivariate least absolute deviation regression method. *Terr. Atmos. Ocean. Sci.*, **18**, 805 – 825.
- Dalcher, A., E. Kalnay, R. N. Hoffman, 1988: Medium Range Lagged Average Forecasts. *Mon. Wea. Rev.*, **116**, 402 – 416.
- Eckel, F. A., and M. K. Walters, 1998: Calibrated probabilistic quantitative precipitation forecasts based on the MRF ensemble. *Wea. Forecasting*, **13**, 1132 – 1147.
- Fujita, T., D. Stenstrud, and D. C. Dowell, 2007: Surface Data Assimilation using an ensemble filter approach with initial condition and model physics uncertainties. *Mon. Wea. Rev.*, **135**, 1846 – 1868.
- Glahn, H. R., and D. A. Lowry, 1972: The use of model output statistics (MOS) in objective weather forecasting. *J. Appl. Meteor.*, **11**, 1203 – 1211.
- Gourley, J. J., J. Zhang, R. A. Maddox, C. M. Calvert, and K. W. Howard, 2001: A real-time precipitation monitoring algorithm-Quantitative Precipitation Estimation Using Multiple Sensors (QPE-SUMS). Preprints, Symp. on Precipitation Extremes: Prediction, Impacts, and Responses, Albuquerque, NM, *Amer. Meteor. Soc.*, 57 – 60.
- Hamill, T. M., 1997: Reliability diagrams for multicategory probabilistic forecasts. *Wea. Forecasting*, **12**, 736 – 741.
- ___, 2001: Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Wea. Rev.*, **129**, 550 – 560.
- ___, and J. Juras, 2006: Measuring forecast skill: Is it real skill or is it

- the varying climatology? *Quart. J. Roy. Meteor. Soc.*, **132**, 2905 – 2923.
- Hartmann, D. L., R. Buizza, T. N. Palmer, 1995: Singular Vectors: The Effect of Spatial Scale on Linear Growth of Disturbances. *J. Atmos. Sci.*, **52**, 3885 – 3894.
- Hsiao, L. F., M. J. Yang, C. S., Lee, H. C., Kuo, D. S., Shih, C. C., Tsai, C. J., Wang, L. Y., Chang, Y. C., Chen, L. Feng, J. S., Hong, C. T., Fong, D. S., Chen, T. C., Yeh, C. Y., Huang, W. D., Guo, and G. F. Lin, 2013: Ensemble forecasting of typhoon rainfall and floods over a mountainous watershed in Taiwan. *J. Hydrology*, **506**, 55 – 68.
- Hsu, W.-R., and A. H. Murphy, 1986: The attributes diagram: A geometrical framework for assessing the quality of probability forecasts. *Int. J. Forecasting*, **2**, 285 – 293.
- Jolliffe, I. T., and D. B. Stephenson, 2003: *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. Wiley, 254 pp.
- Kalnay, E., and A. Dalcher, 1987: Forecasting forecast skill. *Mon. Wea. Rev.*, **115**, 349 – 356.
- Kim, H. S., C.-H. Ho, P.-S. Chu, and J.-H. Kim, 2010: Seasonal prediction of summertime tropical cyclone activity over the East China Sea using the least absolute deviation regression and the Poisson regression. *Int. J. Climatol.*, **30**, 210 – 219.
- Lang, S. T. K., M., Leutbecher, and S. C., Jones, 2012: Impact of perturbation methods in the ECMWF ensemble prediction system on tropical cyclone forecasts. *Quart. J. Roy. Meteor. Soc.*, **138**, 2030 – 2046.
- Larson, H. J., 1982: *Introduction to Probability Theory and Statistical Inference*. 3rd ed. John Wiley and Sons, 638 pp.
- Li, J.-S., and J. -S., Hong, 2011: The study of regional ensemble forecast: Physical perturbations (in Chinese with an English abstract). *Atmos. Sci.*, **39**, 95 – 115.
- _____, and J. -S., Hong, 2014: The Study of Regional Ensemble Forecast: Evaluation for the performance of perturbed methods (in Chinese with an English abstract). *Atmos. Sci.*, **42**, 153 – 179.
- Lu, C., H. Yuan, B. Schwartz, and S. Benjamin, 2007: Short-range forecast using time-lagged

- ensembles. *Wea. Forecasting*, **22**, 580 – 595.
- Mann, H. B., 1945: Non-parametric test against trends. *Econometrica*, **13**, 245 – 259.
- Mason, S. J., and N. E. Graham, 1999: Conditional probabilities, relative operating characteristics, and relative operating levels. *Wea. Forecasting*, **14**, 713 – 725.
- Mass, C. F., 2003: IFPS and the future of the National Weather Service. *Wea. Forecasting*, **18**, 75 – 79.
- Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis, 1996: The ECMWF ensemble system: Methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73 – 199.
- Mullen, S. L., and R. Buizza, 2004: Calibration of probabilistic precipitation forecasts from the ECMWF EPS by an artificial neural network. Preprints, *17th Conf. on Probability and Statistics in the Atmospheric Sciences*, Seattle, WA, *Amer. Meteor. Soc.*, J5.6.
- Murphy, A. H., 1977: The value of climatological, categorical and probabilistic forecasts in the cost-loss ratio situation. *Mon. Wea. Rev.*, **105**, 803 – 816.
- ___, 1993: What is a good forecast? An essay on the nature of goodness in weather forecasting. *Wea. Forecasting*, **8**, 281 – 293.
- Norton, C., P.-S. Chu, and T.A. Schroeder, 2011: Estimating changes in future heavy rainfall events for Oahu, Hawaii: A statistical downscaling approach. *J. Geophys. Res. (Atmospheres)*, **116**, D17110, doi:10.1029/2011JD015641.
- Su, S. H., H. C. Kuo, L. H. Hsu, and Y. T. Yang, 2012: Temporal and Spatial Characteristics of Typhoon Extreme Rainfall in Taiwan. *J. Meteor. Soc. Japan*, **90**, 721 – 736.
- Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317 – 2330.
- ___, and E. Kalnay, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297 – 3319.
- Wei, M., Z., Toth, R., Wobus, and Y., Zhu, 2008: Initial perturbations based on the ensemble transform (ET) technique in the NCEP global

- operational forecast system. *Tellus A*, **60**, 62 – 79.
- Whitaker, J. S., and A. F. Loughe, 1998: The relationship between ensemble spread and ensemble mean skill. *Mon. Wea. Rev.*, **126**, 3292 – 3302.
- Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. Academic Press, 704 pp.
- Yeh, T. C., and R. L. Elsberry, 1993: Interaction of typhoons with the Taiwan orography. Part I: Upstream track deflections. *Mon. Wea. Rev.*, **121**, 3193 – 3212.
- Yuan, H., X. Gao, S. L. Mullen, S. Sorooshian, J. Du, and H. H. Juang, 2007: Calibration of probabilistic quantitative precipitation forecasts with an artificial neural network. *Wea. Forecasting*, **22**, 1287 – 1303.
- _____, J. A. McGinley, P. J. Schultz, C. J. Anderson, and C. Lu, 2008: Short-range precipitation forecasts from time-lagged multimodel ensembles during the HMT-West-2006 campaign. *J. Hydrometeor.*, **9**, 477 – 491.
- Zhu, Y., 2005: Ensemble forecast: A new approach to uncertainty and predictability. *Adv. Atmos. Sci.*, **22** (6), 781 – 788.

Evaluation, Calibration and Economic Value Analysis of the Probabilistic Precipitation Forecasts from WRF Ensemble Prediction System in Taiwan Area- Part 1: Forecast Evaluation

**Hui-Ling Chang¹, Kuan-Ju Chen¹, Chia-Jung Wu¹, Tsorung Wang¹,
Jing-Shan Hong¹, and Shu-Chih Yang²**

1. Central Weather Bureau, Taiwan

2. Department of Atmospheric Sciences, National Central University, Taiwan

(manuscript received 05 March 2018 ; in final form 04 May 2018)

Abstract

This study aims to develop the probabilistic quantitative precipitation forecasts (PQPFs) associated with typhoons by utilizing the WRF ensemble prediction system (WEPS) developed by Central Weather Bureau (CWB). The ultimate goal is to provide more valuable precipitation forecast products associated with typhoons based on the WEPS and help users optimize their decision making and obtain the maximum economic value by using the PQPFs. The research results will be displayed in three articles, including evaluation, calibration, and economic value analysis.

In this paper (part I – Evaluation of PQPFs), we focused on the evaluation of the quality of WEPS, generated PQPFs from WEPS, and verified the performance of PQPFs in details. The ensemble spread can well represent the forecast uncertainties. The PQPFs have good discriminating ability but also obvious biases. The reliability and discrimination decrease with increasing precipitation thresholds.

Verification results from different areas show that the reliability and discrimination over land areas surpass those over adjacent sea areas. This is related to the characteristic of radar quantitative precipitation estimation (QPE) used as ground truth. In addition, the reliability over plain areas is better than that over mountain areas, but the discrimination over mountain areas excels that over plain areas. Taken as a whole, the forecast skill over mountain areas surpasses that over plain areas. This is because the terrain-locking effect leads to higher predictability of typhoon rainfall

over windward slopes than over plain areas. Therefore, the discrimination and forecast skill of WEPS over mountain areas are better. The non-parametric Mann-Whitney test indicated that the greater forecast ability over mountain areas than over plain areas is statistically significant.

Key word : probabilistic quantitative precipitation forecasts (PQPFs), reliability, discrimination, forecast skill

doi: 10.3966/025400022018034601003