

細探公正預兆得分 (Equitable Threat Score) 校驗 小區域定量降水預報之可用度

王重傑 張智昇
國立臺灣師範大學地球科學系

摘要

為客觀了解數值模式預報結果之優劣，其定量降水預報 (quantitative precipitation forecast, QPF) 表現往往是被優先探討的環節。在多種傳統技術得分評估方法中，公正預兆得分 (equitable threat score, ETS) 亦為常用者之一。然而，細探其內涵，卻發現當應用於小區域 (例如：臺灣) 時，雖然模式預報與觀測接近，但因不同降雨區 (觀測、預報、命中) 佔整個校驗區達相當比例，使得計算的隨機值過大 (高估)，導致在同一降水事件中，預兆得分 (threat score, TS) 與偏倚得分 (bias score, BS) 結果優異，但ETS卻偏低，無法客觀反應模式實際的預報能力。

本文以2009年莫拉克颱風之24小時累積降水量為例，主觀判斷可見模式之QPF結果有高度的參考價值，不同降水閾值之TS與BS得分亦相當優良，但ETS卻完全無法反映類似的預報品質；而在將ETS內之隨機值利用整個模式區域資料計算修正後，新ETS值已可相當合理呈現不同降水閾值之QPF能力，故本文建議校驗臺灣地區之QPF應採此類修正之ETS為之。

關鍵字：定量降水預報，模式校驗，公正預兆得分

一、前言

降水的模擬結果往往是一項評估、校驗數值模式預報品質優劣的重要環節。除了需與人為主觀研判相符外，亦需能忠實反應模式之真正的表現 (Davis et al. 2006)，傳統上為了能客觀掌握數值模式之定量降水預報 (quantitative precipitation forecast, QPF) 能力，常使用多種技術得分 (表1) 進行評估，如預兆得分 (threat score, TS)、偏倚得分 (bias score, BS) 及公正預兆得分 (equitable threat score, ETS) 等 (Anthes 1983; Schaefer 1990; Mesinger and Black 1992; Wilks 1995; Ebert et al. 2003a、b; Mason 2003)，其相關定義如下 (Schaefer 1990; Wilks 1995)：

$$TS = H/(O+F-H) \quad (1)$$

$$ETS = (H-R)/(O+F-H-R) \quad (2)$$

$$R = (O/N) \times (F/N) \times N = F \times (O/N) = O \times (F/N) \quad (3)$$

$$BS = F/O \quad (4)$$

其中， O 為觀測雨量、 F 為模式預報雨量、 H 則為 O 與 F 兩者之交集、 N 為校驗目標區域，以及 R 為隨機預報。TS值介於0~1間，若TS值愈趨近1，表示預報結果愈佳。而ETS與TS不同之處僅在於ETS較TS多考量了隨機預報的部分 (如(2)式)。

近年來數值模式技術與電腦運算能力已有明顯進步，若能搭配適當的高速平行計算資源，已可嘗試進行較高解析度的日常例行性預報作業，但如此一來，在使用上述多種評估方法過程中卻已可能出現弔詭，對高解析度模擬結果產生所謂的「雙重懲罰 (double penalty)」，因而所得到的評估結果亦不公正 (Ebert and McBride 2000)，Wang (2014) 指出當不同降雨區 (觀測、預報、命中) 佔整個校驗區達相當比例時，此時ETS所計算之隨機值 (R) 過大 (高估) 將導致ETS以非線性趨勢降低，使得ETS偏低無法客觀、合理反應模式預報能力。

因此本文嘗試探討，當在評估相對較小區域 (例百公里 (例如：臺灣地區) 之模式QPF時，使用ETS方法過程可能發生的問題與其改善方法。

二、ETS於較小區域之評估結果— 以莫拉克颱風降水為例

本研究所舉例子為利用美國「國家環境預測中心 (National Centers for Environmental Prediction, NCEP)」全球預報系統 (Global Forecast System, GFS) 之 $1^\circ \times 1^\circ$ 、垂直向26層之即時分析與預報場，內插至雲解析風暴模式 (Cloud-resolving Storm Simulator; Tsuboki and Sakakibara 2002、2007) 格點，所得之2009年莫拉克颱風即時預報結果，相關模式設定與模擬結果請參見Wang et al. (2012、2013)。

以2009年8月8日24小時累積降雨量為例 (圖1)，主觀上已可被認為是一次成功的降水預報，不論是降水的極值與空間分布，模擬結果均相當接近於觀測值。以TS值與BS值綜合評估，在 ≤ 250 mm以下各閾值均呈現相當不錯之技術得分，頗具有預報參考價值；但若同時採以ETS技術進行客觀評估 (結果如圖2)，卻得到差強人意之技術得分。其中，可發現除了1000 mm以外之各閾值，其TS與ETS間之差值 (D 值； $D = TS - ETS$) 均較明顯，此乃因為此時僅取臺灣陸地上之模式格點作為計算樣本，致使 O/N (事件發生機率) 與 F/N (模式樣本率) 較大，於是 R 值屬於過度被放大；另一方面，分析BS值可看出，同樣為BS趨近於1，但在 < 10 mm之各閾值的 D 值約等於1，但於100 ~ 500 mm間之 D 值卻介於0.1 ~ 0.3，於是在此情形下，此一信息可能讓使用者 (預報員) 易失去使用此筆客觀資料的信心度，殊為可惜。

三、ETS之修正使用方式

重新檢視ETS關係式 ((2)與(3)式)，可發現ETS值係由 O 、 F 、 H ，以及 N 等數值所決定，與各個不同降水閾值並無任何關聯，因此上述ETS結果並非僅見於本文之莫拉克極端降水個案，而是此種技術得分偏低之誤差亦常見於其他降水事件中。由 R 值之定義出發，可得知影響 R 值大小的主因為不同降水區 (O 、 F 及 H) 與分析區域 (N) 之面積比，即 O/N 、 F/N 及 H/N 。因此，假定給予 O/N 為50 ~ 100%，則 F/N 與TS、BS、ETS及 $D = TS - ETS$ 之相關結果如圖3所示。可發現到，當 O/N 愈趨近於100%，ETS以非線性趨勢呈現預報不足的得分情形，其 D 值亦明顯增大；以 $O/N = 50\%$ 為例，當 $F/N = 10\%$ 時， D 值已高達0.1，顯見不適當的隨機值 (R) 可能於計算過程中被過度放大 (高估)，致使ETS邊結果偏低。由此可見如何決定 R 值的母體樣本數，將直接影響到ETS最終的大小。而經修正後，在不同比例之 O/N 值情形下，可見ETS (圖3a-3c中之空心圓點曲線)，已可明顯改善原先之過度

預報與預報不足之缺點。因此，本文嘗試將模式部分的 R 值計算方式修正為如(5)式所示：

$$R = O \times F_M / M \quad (5)$$

其中， M 為模式模擬區內總格點數 (如圖4)、 F_M 為模擬區內達不同降水閾值之格點數。此定義與僅擷取模擬區內特定小區域作為 N 值之內涵有所差異。

結果可發現當 R 值改以(5)式計算後所得之ETS值，在莫拉克實際預報案例中 (圖2c之ETSc)，由於達不同降水閾值之預報降水區與模擬區之面積比已大幅降低，使得 R 值不至於被高估，因此計算所得之ETS值已可合理、公正反應模式QPF產品之品質；

四、討論與總結

客觀上為了彌補TS與BS最後未考慮校驗區大小之因素而可能不夠周全之缺點，因而延伸出以納入隨機命中因素的ETS評估方法，然而以實際降水事件細探其評估方式，發現仍似存在著弔詭之處，當降雨區與校驗區面積相近時 (如降水閾值偏向小雨時)，甚或因地形鎖定效應之颱風環流降水結果，便易出現TS與BS相當優異，但ETS卻無法呈現應有之預報技術；又或是欲評估模式對致災性豪雨之QPF優劣，卻因為分析區域母體樣本數選取問題，無意中使得高降水閾值的ETS得分過於偏低。

本文藉由實際案例的分析，針對ETS之合理使用方式作一探討，嘗試以修正模擬區樣本數的方式 (F/N)，以得到較為合理之 R 值 (當然，亦可著手修正 O/N ，此時 O 值應為氣候值，而非直接採用當次降水事件)，期盼能對合理評估數值模式之模擬結果 (或QPF) 有所助益。

參考文獻

- Anthes, R. A., 1983: Regional model of the atmosphere in middle latitudes. *Mon. Wea. Rev.*, **111**, 1306–1335.
- Davis, C., B. Brown, and R. Bullock, 2006: Object-based verification of precipitation forecasts. Part I: Methodology and application to mesoscale rain areas. *Mon. Wea. Rev.*, **134**, 1772–1784.
- Ebert, E. E., and J. L. McBride, 2000: Verification of precipitation in weather systems: Determination of systematic errors. *J. Hydrol.*, **239**, 179–202.
- Ebert, E. E., U. Damrath, W. Wergen, and M. E. Baldwin, 2003a: The WGNE assessment of short-term quantitative precipitation forecasts (QPFs) from operational numerical weather

prediction models. *Bull. Amer. Meteor. Soc.*, **84**, 481–492.

Ebert, E., U. Damrath, W. Wergen, and M. E. Baldwin, 2003b: Supplement to the WGNE assessment of short-term quantitative precipitation forecasts. *Bull. Amer. Meteor. Soc.*, **84**, 492.

Mason, I. B., 2003: Binary events. *Forecast Verification--- A Practitioner's guide in Atmospheric Science*, I. T. Jolliffe and D. B. Stephenson, Eds., Wiley and Sons, 37–76.

Mesinger, F., and T. L. Black, 1992: On the impact on forecast accuracy of the step-mountain (eta) vs. sigma coordinate. *Meteor. Atmos. Phys.*, **50**, 47–60.

Schaefer, J. T., 1990: The critical success index as an indicator of warning skill. *Wea. Forecasting*, **5**, 570–575.

Tsuboki, K., and A. Sakakibara, 2002: Large-scale parallel computing of cloud resolving storm simulator. *High Performance Computing*, Springer, H. P. Zima et al. Eds., 243–259.

Tsuboki, K., and A. Sakakibara, 2007: Numerical Prediction of High-Impact Weather Systems. *The Textbook for Seventeenth IHP Training Course in 2007*. HyARC, Nagoya University and UNESCO, 273 pp.

Wang, C.-C., H.-C. Kuo, Y.-H. Chen, H.-L. Huang, C.-H. Chung, and K. Tsuboki, 2012: Effects of asymmetric latent heating on typhoon movement

crossing Taiwan: The case of Morakot (2009) with extreme rainfall. *J. Atmos. Sci.*, **69**, 3172–3196.

Wang, C.-C., H.-C. Kuo, T.-C. Yeh, C.-H. Chung, Y.-H. Chen, S.-Y. Huang, Y.-W. Wang, and C.-H. Liu, 2013: High-resolution quantitative precipitation forecasts and simulations by the Cloud-Resolving Storm Simulator (CReSS) for Typhoon Morakot (2009). *J. Hydrol.*, **506**, 26–41.

Wang, C.-C., 2014: On the calculation and correction of equitable threat score for model quantitative precipitation forecasts for small verification areas: The example of Taiwan. *Wea. Forecasting*, **29**, 788–798.

Wilks, D. S., 1995: *Statistical methods in the atmospheric sciences*. Academic Press, 467 pp.

表1 傳統評估模式QPF技術得分2×2列聯表。

		觀測		總和
		是	否	
預報	是	H (命中)	$F-H$ (誤報)	F
	否	$O-H$ (未命中)	$N-(O+F-H)$ (正確預報)	$N-F$
總和		O	$N-O$	N

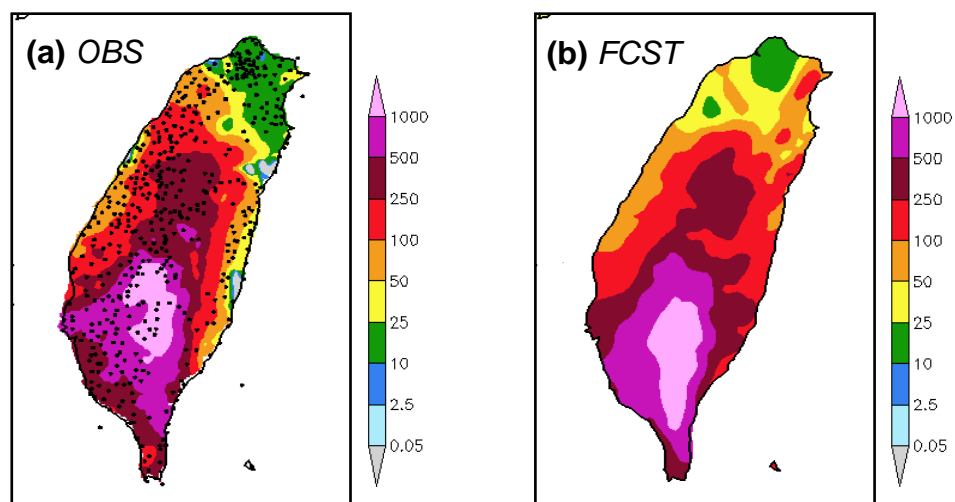


圖1 2009年8月8日0000~2400累積降水量(單位: mm)。(a)觀測值,黑點為雨量站與(b)為CReSS模式(網格間距4公里)之第24~48小時24小時累積預報量(模式初始時間為8月7日0000 UTC)。

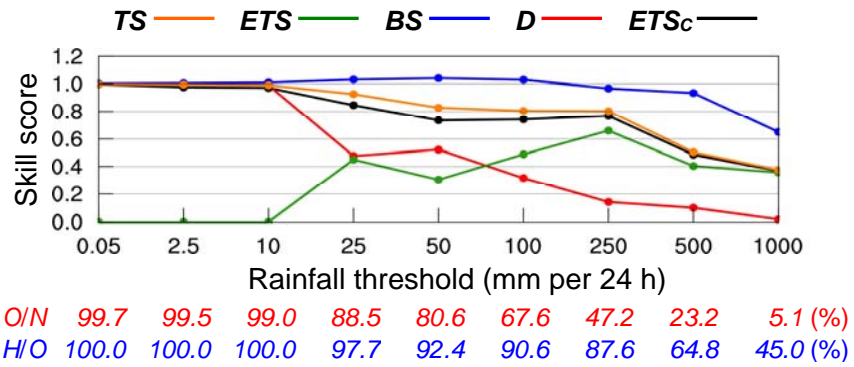


圖2 校驗CReSS模式預報2009年8月8日0000~2400 UTC之TS、ETS、BS、D及ETS_c (R值之母體樣本數為模式模擬區域總格點數)於不同降水量閾值(0.05~1000 mm)之得分, O/N值與H/O值亦列出以供參考。

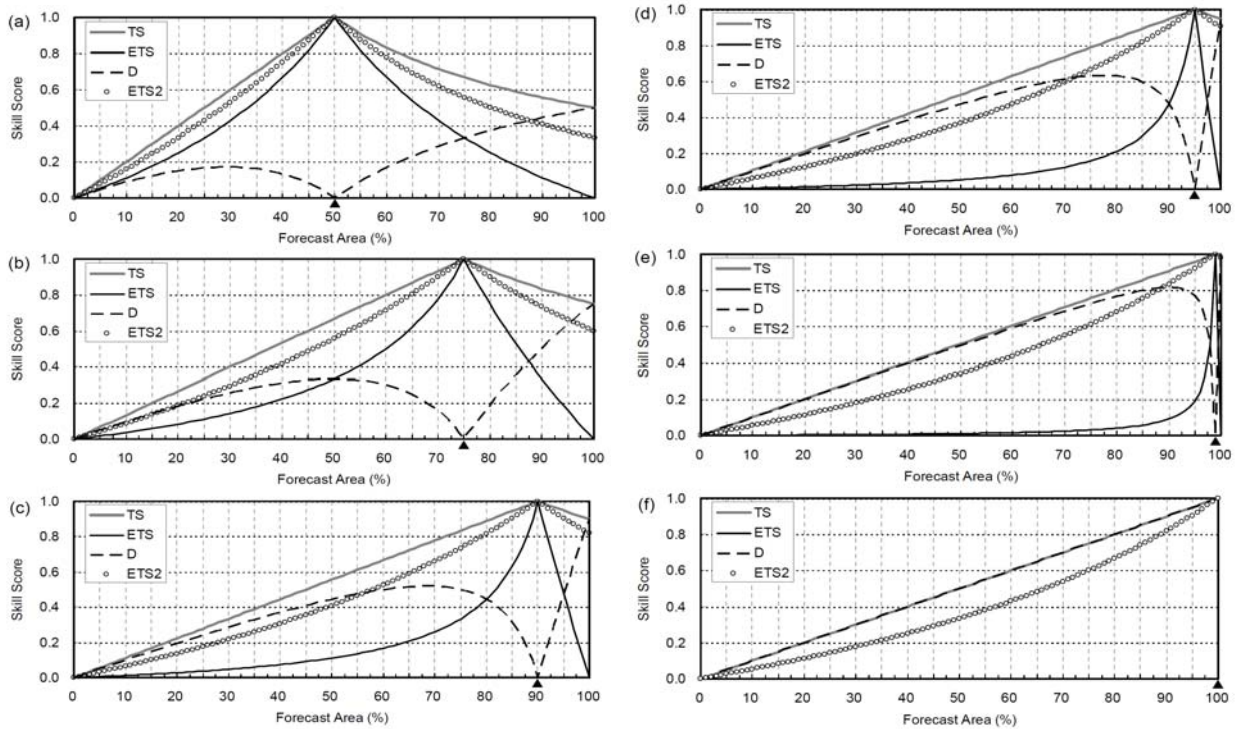


圖3 取觀測降水區(O/N值)為(a) 50%、(b) 75%、(c) 90%、(d) 95%、(e) 99%及(f) 100%時, 其TS(灰線)、ETS(黑線)、兩者之差值D(=TS-ETS; 虛線), 以及經R值修正後之ETS2(空心圓)分布圖。

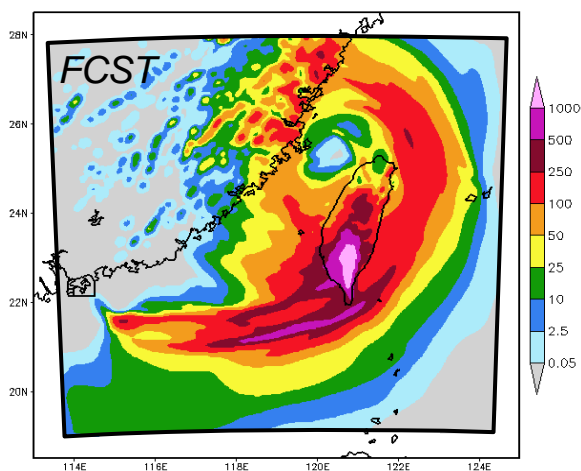


圖4 CReSS 模擬區(4公里網格間距)與其2009年8月8日之24小時累積降水分布圖。