

# 台灣週累積降雨分型機率預報

陳昫靖<sup>1</sup> 張惠玲<sup>2</sup> 陳孟詩<sup>3</sup> 羅存文<sup>1</sup> 周柿均<sup>1</sup> 朱心宇<sup>4</sup>  
科技中心<sup>1</sup> 資訊中心<sup>2</sup> 第三組<sup>3</sup> 預報中心<sup>4</sup>

## 摘要

由於全球模式的網格解析度較低，在台灣這種地形複雜的地區，較難以掌握受地形影響的局部降雨，無法提供正確的降雨分佈。本研究應用SubX計畫之EMC GEFS模式預報之環流場，透過統計後處理方法可連結模式預報之大尺度環流與測站降雨。其結果表明，若模式預報正確之大尺度環流趨勢，透過此方法即可預報出較符合真實的降雨分布。

此外，隨著預報時間拉長，預報誤差迅速增加，模式難以提供精確的大尺度環流資訊。在此透過自組織映射圖網路(Self-Organizing Map, SOM)將各個季節降雨進行分類，並提供不同降雨型態之機率預報，藉以呈現出預報之不確定性。

透過上述流程，能讓使用者了解各個季節在不同大尺度環流條件台灣的降雨分佈，且能了解各種雨型發生的可能性。

關鍵字：降雨分型、統計後處理

## 一、前言

隨著全球數值預報模式之預報能力提升，模式預報到第二週對於大尺度系統之趨勢變化仍有不錯的掌握。陳等(2019)透過統計後處理連結數值模式預報之大尺度資訊與台灣測站週累積降雨。其結果顯示該方法到了第二週對於台灣測站降雨仍可提供具預報參考價值之三分類機率預報，這也代表能透過模式預報之大尺度資訊對於測站降雨變化趨勢進行推估。

相對於三分類機率預報，定量降雨預報對於農業以及水利等單位有較高的應用價值，因此本文參考陳等(2019)之方法，選用SubX(Subseasonal Experiment)計畫下之EMC-GEFS模式，透過統計後處理之方式，將大尺度預報因子轉換成定量降雨預報。此外，為了顯現出預報之不確定性，本文透過Kohonen(1982)提出的自我組織映射(Self Organizing Map, SOM)針對台灣地區不同季節之週累積雨量分布分類成3至4個類別，並預報各個類別發生之機率。透過上述流程除了整合眾多成員提供的資訊外，也提供使用者各個季節常見的降雨分布型態。本文將以春季為例，詳述整個統計後處理之流程與預報表現評估。

## 二、資料來源與研究方法

### (一) 使用資料

在此使用EMC-GEFS模式歷史預報資料(1999-2016年2至4月)做為訓練期，即時預報資料(2017-2020年2至4月)為校驗期之預報1-3週週累積降雨預報資料。其網格解析度為 $1^{\circ} \times 1^{\circ}$ ，歷史預報有11個成員而即時預報則有21個成員。觀測資料則使用318個測站對應預報時段之降雨資料，其測站分布如圖(一)所示。

### (二) 建模流程

在此採用前後22天共45天的移動窗區來選取建模資料。舉例來說，當2020/02/01做預報時，會將1999-2016年初使預報時間位於之01/10-02/23之間所有預報個案都納入作為建模資料。因在季節轉換季，例如在4月底時轉難決定到底該用梅雨季還是春季的資料建模。此建模流程因都抓取每年該時段附近日期之個案作為建模資料，可避免跨季節的問題。

#### 1. Logistic Regression :

$$P(Y=1 | X) = \frac{\exp(a+bx)}{1 + \exp(a+bx)} \text{----- 式(一)}$$

當Y=1時代表事件發生，Y=0時代表事件未發生，X為預報因子。P(Y=1 | X)即代表當預報因子為X時，事件發生之機率。與線性回歸不同，羅吉斯回歸透過最大似估計法去求得參數a以及b使得 $\prod_i^n p_i^{y_i} (1 - p_i)^{1-y_i}$ 達到最大值。其中，其中當Y<sub>i</sub>=1其得分為p<sub>i</sub>，當Y<sub>i</sub>=0其得分為1-p<sub>i</sub>。

#### 2. Extend Logistic Regression(XLR) :

$P(Y=1 | X) = (\exp(a+bx+cQ)) / (1+\exp(a+bx+cQ))$  —— 式(二)  
式(二)與式(一)差別在於，式(二)多了cQ這項。其中Q為百分位，變數c則為待求之係數。除了本身的預報因子外，又額外加入了百分位做為預報因子。此方法的優點在於所有降雨門檻只需要建立一條迴歸方程式，並且保證了預報機率隨著事件門檻的提高為單調遞減。

### 3. 預報因子選取

在此選用三種預報因子分別為

1. 系集平均之測站鄰近四格點模式降雨內插置格點上
2. 系集平均之經度119-123 緯度21-26範圍內模式降雨取EOF前三個mode(PC1-PC3) 迴歸置測站降雨。
3. 系集平均之經度110-140 緯度10-40範圍內模式降雨取EOF前三個mode(PC1-PC3) 迴歸置測站降雨。

上述三種預報因子分別代表模式預報不同尺度之降雨資訊。因為模式預報降雨能力隨著預報時段拉長其預報誤差會逐漸增加。此時模式降雨預報之小尺度資訊會越來越不可信，又因模式的地形解析度不夠高，導致模式模擬不出一些地形造成的局部降雨，因此鄰近格點降雨未必可信，尤其預報到第二週以後抓出大尺度降雨訊號作為預報因子就更有其必要性。

實際預報會從上面三個取一組預報因子，選取的基準為透過建模之歷史個案找出三個預報因子，接著計算出與所有測站降雨之相關性。若是與第三組預報因子相關性最高的測站數最多，則該次預報之所有測站都會選取第三組預報因子來建模並預報。實作經驗中，預報第一週較常選到第一或第二組尺度較小之預報因子，到了第二週後則選取到第三組(尺度較大的)機會較高。

### 4. 測站降雨

如圖(二)所示，各百分位門檻透過Bias趨近於1(預報事件發生之次數接近觀測事件發生之次數)的方法找出該百分位之最佳機率門檻。透過比對每個百分位門檻預報之機率是否超越該百分位之最佳機率門檻，來決定降雨百分位。以此圖為例，在第0百分位預報機率大於最佳機率門檻，因此降雨大於第0百分位，在第5百分位預報機率仍大於最佳機率門檻，因此預報降雨大於第5百分位，依此類推。一直到預報第50百分位預報機率才小於最佳機率門檻。因此這次預報雨量應大於第45且小於第50百分位。透過歷史資料比對，第45百分位對應實際雨量值為38.5，而第50百分位對應之觀測值

為45.4。因此這次預報雨量值介於38.5-45.4之間，在此就會取38.5當作本次定量降雨預報。此圖範例為了說明方便，取的預報百分位間隔較大，實際預報會將上述流程由間隔百分之五改為間隔百分之一，由於間隔較密，各個百分位對應之降雨量差異也較小，如上述直接取45百分位對應之雨量值(38.5mm)作為定量降雨預報影響也較小。對每個測站重複進行上述步驟，即可得到台灣地區之預報降雨分布。至此步驟得到的降雨預報在此暫時稱作統計後處理降雨預報。

系集平均之預報因子透過XLR可以得到一個完整且離散程度較為合理的累積機率分佈(CDF)，為了保有此CDF的離散度，在此將依等百分位重新抽取21個成員(ex.抽取累積機率為0/20,1/20,2/20,...,19/20,20/20對應之21個觀測百分位)。得到預報之觀測百分位之後，再透過歷史預報資料對比得到觀測百分位相應之雨量值，即可得到21個週累積降雨預報。重新取樣後會打亂各個成員的降雨分佈，因此須找回原本的空間相依性。在此先將各個成員依照統計後處理預報得到之雨量大小排名，在依照排名將重新抽取之不同百分位雨量填入該成員之預報(ex.將預報第50百分位對應之雨量值取代原本預報雨量大小的11名的成員預報之雨量值)。透過上述方法即可保留EXL提供之離散度且找回各成員之空間分佈。經過重新抽樣後的結果在此暫時稱為統計後處理+重新抽樣。

### 5. 雨量分型機率預報

首先找出春季常出現的降雨型態。圖(三)為透過SOM將春季之週累積雨量分佈分類為3個類別，及各個類別對應之850百帕風場及大尺度降雨合成圖。其分類結果大致可區分為

1. 台灣受高壓出海影響，只有東北部迎風面有地形降雨。
2. 台灣受鋒面影響，但鋒面強度較弱，台灣降雨主要集中在中部以北地區，另外在中北部山區以及東北部迎鋒面有較顯著的降雨。
3. 台灣受到鋒面影響，除了鋒面強度較強以外，還搭配南風提供之水氣，造成較強的降雨，且南部地區也有較高的降雨機會。

接著將21個統計後處理+重新抽樣之預報降雨分布透過歐式距離找出各個成員分別和哪個類別最接近，即可得該成員

之預報類別。接著分別計算預報至各個類別之成員數，再除以總成員數，即可得到各個類別之預報機率。

最後將上述流程整理為圖(四)，此流程圖說明如何將數值模式資料轉化為降雨分型機率預報。

### 三、預報表現評估

#### 1. 降雨分布改善

圖(五)分別為二週前預報至20180331-20180406以及預報至20180217-20180223之個案。由圖可以看出，不論台灣地區受到什麼天氣系統影響，模式原始輸出的降雨都集中在台灣東南部，經由統計後處理後可得較正確的分佈，並有機會預報出台灣東北部受地形影響之迎鋒面降雨。

圖(六)，分別為預報第一週至第三週系集平均預報之春季週累積雨量之誤差以及與觀測之 Pattern Correlation 繪製成散佈圖，橫軸為相關性，縱軸為 Mean Absolute Error。當預報結果位於圖的右下角時，代表預報誤差越小以及預報降雨分布跟觀測越接近。由圖可得知各週統計後處理過後紅點位右下角的點數明顯多於黑點，代表經統計後處理後可以修正降雨分佈且相較於模式原始降雨預報，有更高的機會得到更接近觀測的預報。

#### 2. 降雨分型預報&機率預報

在此用 Heidke Skill Score (HSS) 針對統計後處理後系集平均之預報類別進行校驗(Wilks, D. S, 1995)，HSS以各分類之隨機預報命中率作為評分基準，若HSS等於1代表完美預報，當HSS大於0代表預報之命中率優於隨機命中率，若HSS小於0代表預報之命中率低於隨機命中率也代表毫無預報技術。

下表為統計後處理後系集平均預報第一週至第三週之HSS。其顯示出，隨著預報不確定性隨著預報時段拉長越來越大，其預報得分迅速降低，雖然到第三週仍稍有預報技術，但也顯示出隨著預報時間拉長更不能只看系集平均之預報結果。

預報週數	Week1	Week2	Week3
HSS	0.43	0.23	0.04

圖(七)為兩週前預報20190202-20190208之個案，此圖包含三個類別之機率預報、系集平均之預報類別及降雨分佈以及對應觀測之降雨類別及分佈。此個案之系集平均無法預報到正確的降雨類別。但透過降雨分型的機率預報，可以得知第一型降雨的可能性也不低。此結果凸顯出，單看系集平均之決定性預報風險仍高，需透過機率預報掌握所有不確定性才能做出更全面的判斷。

最後圖(八)及圖(九)皆為統計後處理後預報第一至三週降雨分型機率預報之 Reliability Diagram。但圖(八)為統計後處理降雨預報，圖(九)則為統計後處理+重新抽樣後調整離散度之結果。其結果顯示經過重新抽樣後更貼近黑色斜線，代表能提供更可信之機率預報。

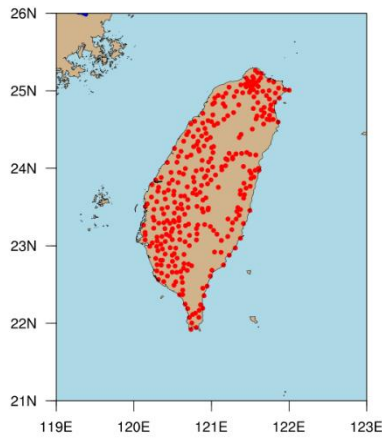
### 四、結論

現今全球模式對於展期預報逐漸能提供具有參考價值之大尺度環流資訊。當模式預報的大尺度變化趨勢正確時，透過上述一系列的流程，就有機會得到較接近觀測的降雨分佈。此外透過降雨分型機率預報不僅能呈現出春季常出現的週累積降雨分佈外，也考慮到模式預報之不確定性，提升預報參考價值。

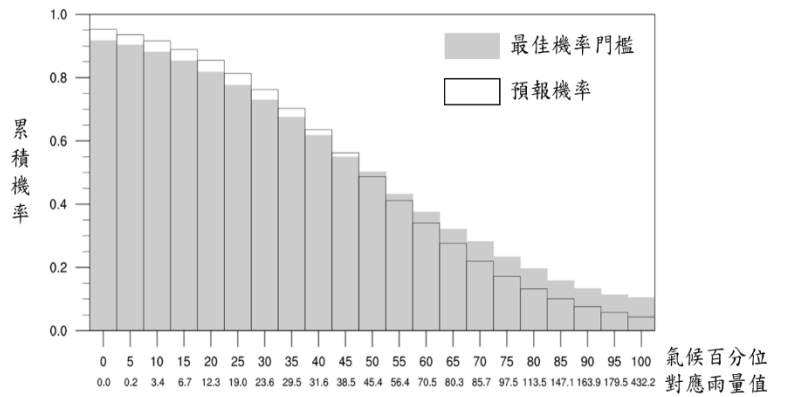
其校驗結果也顯示，以春季來說，到了第三週仍有些許預報技術，且還能提供可信之機率預報，未來將會把此流程套入各個季節測試並評估預報表現。

### 五、參考文獻

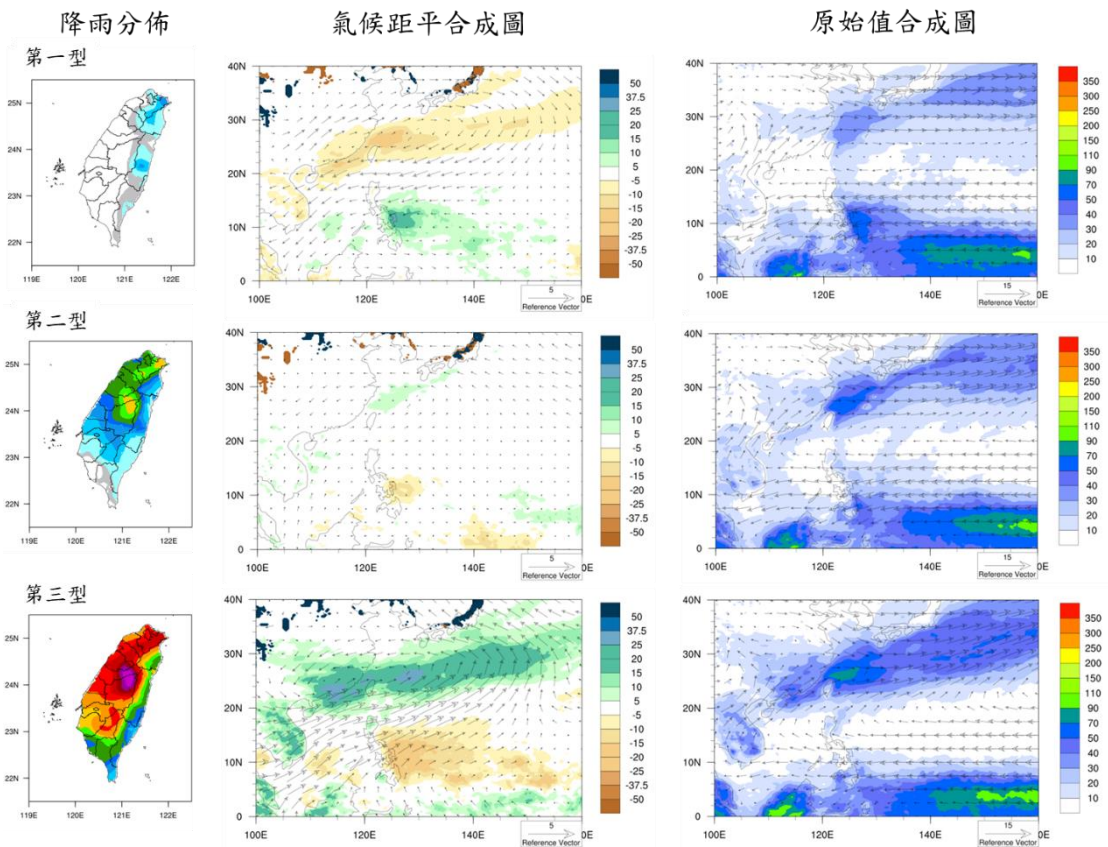
- 陳昀靖, 陳孟詩, 張惠玲, 羅存文, 周柿均, 2019：應用 SubX 模式於 2-4 週測站降雨機率預報, 108 年天氣分析與預報研討會論文全文彙編, A5-20
- Kohonen, T., 1982: Self-organized formation of topologically correct feature maps. *Biol. Cybern.* 43, 59-69.
- Wilks, D. S. 1995: *Statistical Methods in Atmospheric Sciences*. Academic Press, 467p.



圖(一) 測站分佈圖



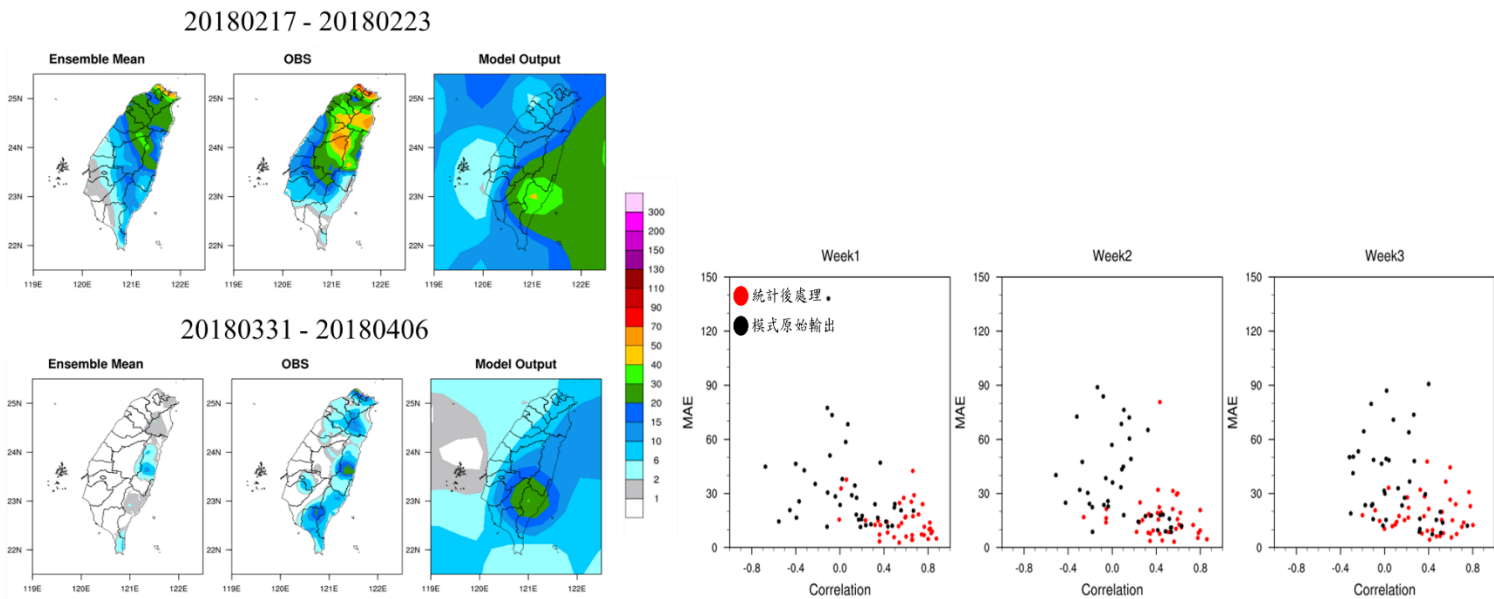
圖(二) XLR得到之CDF及機率預報轉換為定量預報範例



圖(三) 台灣春季週累積雨量分型及對應大尺度850hPa風場及降水合成圖。



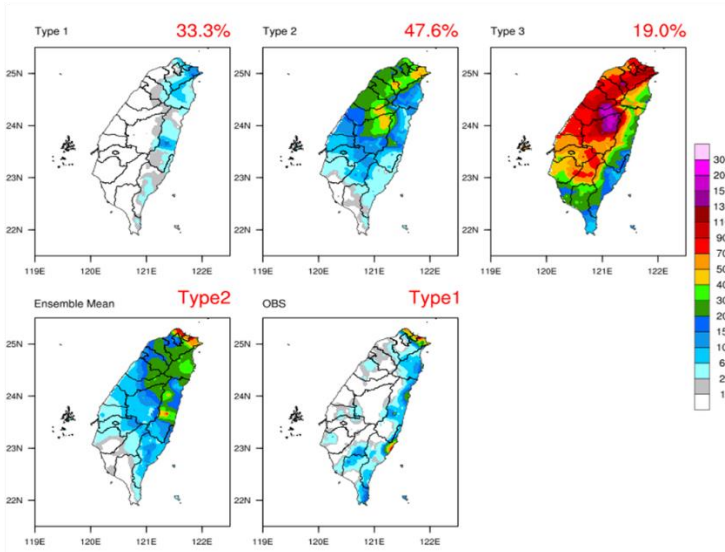
圖(四) 降雨分型機率預報流程圖



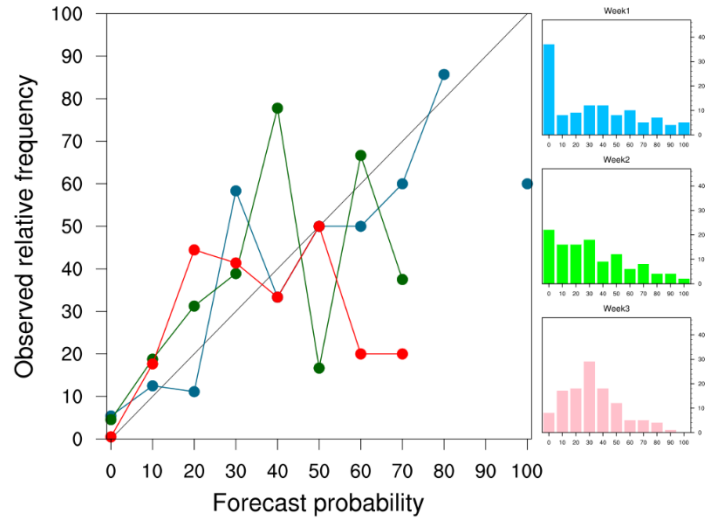
圖(五) 兩週前預報20180217-20180223以及20180331-20180406之週累積降雨。由左至右分別為統計後處理後之系集平均，對應實際觀測以及模式原始輸出之系集平均。

圖(六) 2017-2020春季預報第一至三週所有個案之誤差與空間相關性之散佈圖。紅點為統計後處理之結果，黑點為模式原始輸出之結果。橫軸為空間相關性，縱軸為MAE，越靠近右下角代表空間分布與量值越接近觀測。

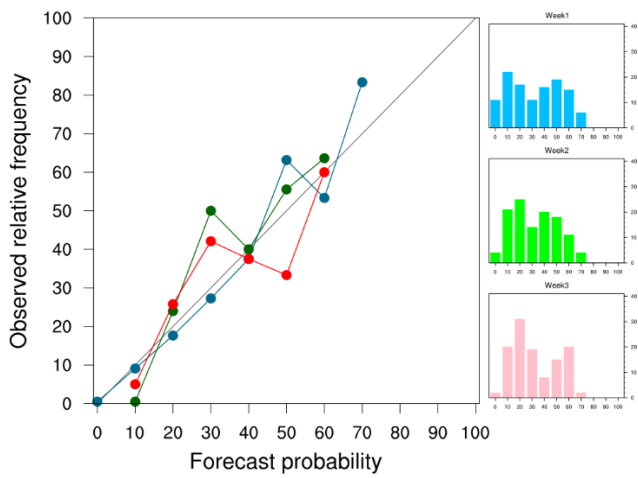
20190202-20190208



圖(七) 兩週前預報至20190202-20190208之個案結果，包含三個降雨分型之機率預報，以及系集平均與對應觀測之降分布及兩型類別。



圖(八) 2017-2020春季預報第一至三週統計後處理後預報降雨分型機率預報之 Reliability Daigram。



圖(九) 同圖(八)，但為統計後處理+重新抽樣之預報結果。