

次季節尺度多變數自迴歸統計預報模式的建置

李永安¹ 彭于珈¹ 李思瑩² 隋中興^{1,2}
大氣物理研究所¹ 水文科學研究所²
國立中央大學

摘要

本研究藉自迴歸的方法，將一組資料構成的預報因子 $Y(t)$ 及被預報因子 $Y(t+\tau)$ ，求二者之間存在著線性關係式 $Y(t+\tau) = AY(t)$ 中的線性傳遞矩陣 A ，發展一套多變數自迴歸統計預報模式，AR1(II)。模式的發展主要是利用主分量分析萃取出預報及被預報資料的主要經驗正交模，濾除雜訊，進而透過矩陣運算求得線性傳遞矩陣。

本模式的預報資料以外逸長波輻射（簡稱OLR）為主，渦度場的資料為輔。預報的區域為有二：大區域（40°E ~ 180°E, 25°S ~ 25°N）與小區域（50°E ~ 140°E, 5°N ~ 40°N）。首先使用單一變數模式做OLR預報驗證。在大區域預報範圍內，五六月期間的第25天預報值，在印度半島、赤道西太平洋及海洋大陸一帶，與觀測值之相關係數可達0.4以上；七八月期間則略為遜色。在小區域預報範圍內，五六月期間的預報能力變化不大，而七八月期間模式對鋒面雨帶之預報能力明顯提高。在臺灣地區的5天（10天）預報與觀測值之相關係數有0.3~0.4（0.15）。我們也使用多變數模式，將NCEP/NCAR R2的850 hPa的渦度場加入初始條件做預報，其預報表現與單一變數模式相當。

除了事後預報的驗證，文中也將AR1(II)與其它模式比較。我們選取兩個動力模式（NCEP CFS及氣象局SMIP系集資料）及三個統計預報的模式（Waliser et al. 1999、Jones et al. 2004、Mo 2001）。AR1(II)與動力模式的比較難分軒輊；至於統計模式的預報驗證，各模式在第25天的預報其實差異不大，但是，AR1(II)模式使用沒有濾波的資料即可達到與其它模式同樣的效果，在某些地區甚至可以更好。

關鍵字：次季節尺度氣候預報、自迴歸統計預報模式、線性傳遞矩陣。

一、前言

東亞季風區夏季受到多重尺度氣候/天氣的相伴影響，要瞭解此複雜氣候系統的時空特徵，並發展預報模式是極具挑戰性的研究。過去的研究指出，在月與季的氣候預報中，統計與動力模式的預報結果相差不多。Barnston et al. (1999) 比較8個動力模式及7個統計模式在1997/98的聖嬰事件和1998年反聖嬰肇始，他們發現，雖然不少模式都預報到1997年北半球的聖嬰現象，但對於它的強度確無法明確的掌握，不論是統計模式或是動力模式都一樣。陳等(2003)也提到：對於亞洲季風系統整體特性的季節預報，統計預報通常具有合理程度的預報能力，特別是在強邊界條件時。至於次季節尺度的氣候預報，比月與季的預報更具挑戰性，統計與動力模式各有其優點與局限。我們考量到氣候動力模式的複雜性，應屬預報單位的研發業務，本研究以統計模式的發展為目標，係考量設計較簡單，易與觀測分析結合，計算的時間可以大為縮短。下文將近幾年發表的季內尺度統計預報模式，做簡短的介紹。

Waliser et al. (1999) 使用 National Oceanic and Atmospheric Administration (NOAA) 的外逸長波輻射 (Outgoing Longwave Radiation, OLR) 的候平均 (pentad) 資料，對時間方面做30~70天的濾波，空間的解析度為5° x 5°。模式設置中，預報因子包含預報時間及其前一候平均的資料 (T_N, T_{N-1})；被預報因子則為預報初始時間的未來資料 ($T_{N+\tau}, \tau = 1\sim 9$)。主要使用的統計方法為奇異值分析 (Singular Value Decomposition, SVD)，分別對不同預報時間的預報及被預報因子做SVD分析，並取其前幾個模。將其預報結果與週期30~70天的資料做相關，在第5~20天的預

報大約可以有0.5~0.9的預報技巧；但是與未濾波資料的相關就降至0.3~0.5的預報能力。其在夏季的預報以印度次大陸、東南亞及海洋大陸有較高預報能力。

Jones et al. (2004) 使用週期20~90天的候平均資料；空間的解析度為5° x 5°。空間方面的濾波除了將2.5° x 2.5°平均成5° x 5°外，並對每一個格點用本身一點及周圍八點做權重平均。模式使用的資料有三，OLR、U200及U850，預報資料重組的主分量是先得自三組資料聯合經驗正交函數 (combined Empirical Orthogonal Function, C-EOF) 的分解的預報，再將此預報值最近的五個候平均資料組成；空間分量則直接取自各變數EOF分析後的特徵向量。結果顯示，夏季時預報準確的地方則是在印度、印尼及東太平洋，這些區域大約可以持續到第20天的預報都還可以有0.5的預報能力。

Mo (2001) 使用的統計方法為奇譜分析 (Singular Spectrum Analysis, SSA)。其時間的解析度為週期10~90天的日平均及候平均資料，空間的解析度為2.5° x 2.5°。作者將美國加州 (118°W ~ 123°W, 32.5°N ~ 40°N) 地區的OLR anomaly (OLRA) 做區域平均，做為模式的指標。SSA分析時所決定的windows數不論是日平均 (M=41) 或候平均 (M=17)，在做預報時，主要考慮的是季內振盪 (intraseasonal oscillation, ISO) 中三個週期的模，分別為40天、22天及18天，並透過文章中的公式(7)做外推的預報。候平均的預報結果顯示，第20天預報能力的區域平均為0.65。

本研究將會介紹一套自迴歸統計預報模式，透過一些試驗找到對模式最有效的設定。AR1(II)模式的介紹請看第三節。第四節為模式的試驗及改進。第五節為模式本身的個案

比較及與其它模式的比較。第六節則總結本文。

二、資料

本模式使用的資料有二，其一為NOAA OLR，其二為National Centers for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) Reanalysis-2 (R2) 850 hPa的高度場。解析度皆為 $2.5^\circ \times 2.5^\circ$ ，時間長度為1979年到2005年的候平均資料，預報時間以夏季(五月到八月)為主。

除了上述資料，我們另外選取兩組動力預報模式的資料與本模式比較。分別為中央氣象局SMIP及NCEP Climate Forecast System (CFS) 的系集資料。SMIP的解析度則是採用T42網格，時間長度從1980年到2004年的季預報資料(一月、四月、七月、十月)，每一季預報七個月。CFS的水平解析度為高斯網格，預報的時間從1981年到2003年，每次預報九個月，詳細的資料說明可參考Saha et. al (2005) 或 <http://cfs.ncep.noaa.gov/>。

三、模式設計

本研究發展之自迴歸模式 (autoregressive model)，簡稱AR(1)。下面介紹本模式的設計。

假設現在有一組資料 Y ，其時間長度為 t ，令 $Y(t)$ 為預報因子，與 $Y(t)$ 間隔 τ 時間的 $Y(t + \tau)$ 為被預報因子，兩者之間存在著線性的關係為 $Y(t + \tau) = AY(t)$ ；而使 $Y(t)$ 及 $Y(t + \tau)$ 得以達到最佳線性關係的係數 A 則是本研究的重點—線性傳遞矩陣 (linear propagator matrix)。

預報及被預報因子的設計如下。假若現在要預報2005年的某一時間 t_1 的 Y ； $Y(t)$ 為1979年到2004年 t_1 前後 N 個時間，也就是每年都取 $t_1 - N$ 到 $t_1 + N$ ，共 $2N + 1$ 個 Y 值； $Y(t + \tau)$ 則為每年 $t_1 - N + 1$ 到 $t_1 + N + 1$ (假設 $\tau = 1$)，最後再取2005年的 $t_1 - N$ 到 $t_1 + N$ 當做 $Y'(t)$ 。以上即是模式的初始資料，接下來說明本模式求 A 與事後預報的步驟。本研究使用 Jackknife (亦稱 Cross-Validation) 的方法做預報，可參考Michaelsen (1987) 及 Yuval (2000) 等文章。

1. 首先對 $Y(t)$ 及 $Y(t + \tau)$ 各別做主分量分析 (Principle Component Analysis, PCA)。對 $Y(t)$ ($Y(t + \tau)$) 而言，分解出一個特徵向量 E_1 (E_2) 及主分量 $Z(t)$ ($Z(t + \tau)$)，其矩陣形式為 $Y(t) = E_1 Z(t)$ ($Y(t + \tau) = E_2 Z(t + \tau)$)。接著將特徵向量 E 投影至 $Y'(t)$ 即得到另一組新的時間序列 $Z'(t)$ ($Z'(t) = E_1^T Y'(t)$)。最後取時間序列 $Z(t)$ 、 $Z(t + \tau)$ 及 $Z'(t)$ 中的前幾個模 (mode)，為接下來求線性傳遞矩陣做前置準備。這一步的目的不只可以濾掉一些雜訊，亦可加快模式的計算時間。
2. 從關係式 $A = Z(t + \tau)Z(t)^{-1}$ ，求線性傳遞矩陣 A 。
3. 預報：將線性傳遞矩陣乘上2005年 t_1 的時間序列 Z' (就是預報時間的前一個時間間隔，假設 $\tau = 1$ 個時間間隔)，以方程式表示可寫成 $pred1 = AZ'(t_1)$ ，此乃第一個預報時間。第二個及其後的預報時間重覆上述1到3的步驟，重新計算線性傳遞矩陣；此外步驟1中的 τ 值將隨預報的時間增加。每一個時間的預報可寫成：
 $pred1 = A_1 Z'(t_1)$ ($A_1 = Z(t + 1)Z(t)^{-1}$)

$$pred2 = A_2 Z'(t_1)$$

$$(A_2 = Z(t + 2)Z(t)^{-1})$$

$$pred3 = A_3 Z'(t_1)$$

$$(A_3 = Z(t + 3)Z(t)^{-1})$$

$$pred4 = A_4 Z'(t_1) \dots \dots \dots$$

4. 驗證：使用 Jackknife 方法做事後預報，即使用所有年份的資料，欲對某一年做預報時，即將那一年的資料去掉，用剩餘年份的資料對那一年做事後預報。最後再求預報值與觀測值之間的相關來做驗證。

四、事後預報實驗與分析

在預報之前，必需先設定好兩個參數讓預報的結果可以有最好的成效，此二參數為(a) 預報及被預報因子每年所選取的時間長度，就是預報時間的前後 N 點，共 $2N + 1$ 點，下面統稱 windows；(b) 初始條件 PCA 分析時欲保留的模數。為了找出較適合的設定值，我們從 $40^\circ E \sim 180^\circ E$ ， $25^\circ S \sim 25^\circ N$ 選取四個地區：(1) 印度半島西岸、(2) 台灣與華南地區、(3) 南海地區、(4) 海洋大陸，比較其區域平均的預報能力在上面兩個參數的設定中何者較好。

比較之後，本研究決定 1) 做為初始條件的預報及被預報因子，其使用的資料時間長度為 windows = 11，也就是預報時間的前後五個候；2) 初始條件在 PCA 分析後保留 10 個模的設定來預報。

單變數預報模式

一開始模式設定的預報區域為 $40^\circ E \sim 180^\circ E$ ， $25^\circ S \sim 25^\circ N$ ；並將四個月的預報分成五六月及七八月兩段做驗證。首先將AR(1)模式的預報結果與相對應的觀測資料求相關係數(後文中稱此為預報驗證)，並與OLR自身延遲時間的相關係數(後文中稱此為持續性預報，persistent)比較。

圖一、圖二是 OLR 的持續性預報及預報驗證的比較，大致來看，它們在第5天的預報能力與持續性預報差不多，但是預報時間越長，其預報能力明顯的較持續性預報好。圖一顯示五六月時，在印度半島、赤道西太平洋及海洋大陸預報結果不錯，在印度地區直到25天預報還有0.6的預報能力，赤道西太平洋也有0.4。從圖二可看到七八月的預報能力不如五六月來得好，在海洋大陸最好可有0.6的預報技巧，第20天的預報在海洋大陸及印度西岸也還有0.4。

雖然上面的預報結果還算不錯，但是，台灣地區的預報結果還是不夠好。而且對台灣地區來說，這個區域太大，因此將範圍縮小到東經90度到140度，北緯5度到40度，結果如圖三(一、三欄)所示。在五六月赤道西太平洋及孟加拉灣的預報很不錯，有些地方甚至可以達到0.6以上，而且此高預報能力可以維持到20~25天。七八月在西北太平洋的預報有不錯的表現，第5~25天的預報都可以模擬出一條雨帶區。仔細觀察台灣地區，在五六月及七八月的5天預報可達0.3~0.4的預報能力，10天預報尚有0.15，之後的預報大約維持在0.1上下。

多變數預報模式

為了更加強台灣地區的預報能力，我們在初始資料中多加入一組變數。其預報及被預報因子的關係式為 $Y(t + \tau) = AY(t) + BX(t)$ ；以矩陣型態表示可寫成

$$Y(t + \tau) = \begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} Y(t) \\ X(t) \end{bmatrix}$$

如此一來現在的線性傳遞矩陣將不在只有一個 A ，而是 $\begin{bmatrix} A & B \end{bmatrix}$ 。

我們透過關係式 $Y(t+\tau) = KX(t)$ 及簡單的統計方法 SVD 來決定加入的變數為何。取 OLR 為 $Y(t+\tau)$ (對照組) 及其他的 NCEP R2 變數 (實驗組) 為 $X(t)$ ，對 K ($K = Y(t+\tau)X(t)^{-1}$) 做 SVD 分析。比較對照組及實驗組的兩個空間分佈及 SVD 解出來的奇異值 (singular value)，以決定最適合的變數。

經過比較及考量，因為渦度場對大氣環流較具有影響力，所以選擇使用 850 hPa 的渦度場與 OLR 配合做為初始條件，進行預報驗證的作業。

圖三 (二、四欄) 為雙重變數 (850 hPa 的渦度場與 OLR 同時做為預報因子的組合) 的預報驗證，與圖三 (一、三欄) 比較，的確有改進，但變化不大，而且時好時壞。對較為明顯的特徵型態 (五六月：赤道西太平洋及孟加拉灣；七八月：西北太平洋的兩帶) 來說，幾乎是不變的。取台灣的區域平均 (117°E ~ 125°E, 20°N ~ 27°N) 來看，雙重變數對五六月的預報能力有較好的改善；但在七八月，單一變數 OLR 的預報能力對雙重變數預報卻是略勝一籌。

五、統計與動力模式的比較

ARI(II)與動力模式的比較

目前我們的統計模式還是無法讓台灣地區在 10 天之後的預報達到有用的程度，這樣的結果與動力模式結果相近，孰優孰劣尚無參考資料。我們選了兩個動力模式做比較，其一是 NCEP CFS，其二為中央氣象局的 SMIP。

從八月份的預報驗證來看 (圖四)，這三個模式對兩帶的模擬都還不錯，但 ARI(II) 的結果較不明顯；SMIP 與 CFS 捉到的兩帶在位置上則有些差異，SMIP 較偏北。此外，ARI(II) 及 CFS 在中國華北的預報能力可達 0.4 以上；而中國華南沿岸的預報則是以 ARI(II) 的單一變數模式及 CFS 表現最好，預報能力可達 0.4 以上。SMIP 不同於另幾個模式，它在青藏高原南側及西太平洋的預報比其它的模式都好。雖然這幾個模式各自在不同的地方有突出的表現，但是對台灣地區的預報卻不是很好，大概只有 ARI(II) 的單一模式有達到 0.2 ~ 0.3 的預報能力。

ARI(II)與其它統計模式的比較

本文所提及的四種統計模式，不論模式的設計為何，最主要都是先用統計方法 (SSA、SVD) 將資料分解成特徵向量及主分量，再使用各自發展的預報模式對主分量做預報，最後再將這個預報出的時間與各變數的空間型態重組，得到最後的預報值。

從 Waliser et al. (1999) 圖 11 中雖然可以在東半球找到較高的預報值，在第 25 天預報還有約 0.7 以上；但在其圖 14 中東半球的預報能力降到只剩 0.3 ~ 0.5。整體而言，Mo (2001) 的第 25 天預報，約為 0.3 ~ 0.7，但大概只在海洋大陸擁有較高的預報技巧。Jones et al. (2004) 的第 20 天預報，在印度洋至西太平洋一帶有 0.6 以上的高預報性，但到第 25 天預報只剩印度洋一帶有 0.4 左右的預報能力。ARI(II) 模式在第 5 天預報雖然只有 0.6 以上的預報能力，但是至少在第 25 天預報時，印度大陸、西太平洋及海洋大陸還可以有 0.4 以上的預報能力。

六、結論

ARI(II) 模式在本預報實驗的最優參數設定有二：第一：做為初始條件的預報及被預報因子，其使用的資料時間長度為 windows = 11，也就是預報時間的前後五個候；第二：初

始條件在 PCA 分析後決定保留 10 個模。

單一變數 OLR 在 90°E ~ 140°E, 25°S ~ 25°N 的預報，不論五六月或七八月，其驗證結果都較持續性預報來得好。五六月在赤道西太平洋、印度半島一帶及海洋大陸的預報表現較佳，七八月則在海洋大陸及印度西岸的預報有不錯的表現。雖然 ARI(II) 模式與 Jones et al. (2004) 夏天的預報結果相當，但其對資料做空間及時間上的濾波，我們則只是單純的使用候平均的資料即可預報出類似的結果；與動力模式的比較也是難分軒輊。因本研究預報的重點為台灣地區，所以曾進一步將範圍縮小到 90°E ~ 140°E, 5°N ~ 40°N。縮小範圍後，五六月的預報在赤道西太平洋及孟加拉灣依然有不錯的成績；七八月則是對雨帶的預報上有進一步的改進。此外，台灣地區在五六月及七八月的第 5 天預報都可達到 0.3 ~ 0.4 的預報；10 天預報也可達 0.15。對台灣地區的預報而言，七八月前三個候較五六月來得好，這可能是因為七八月對鋒面雨帶的掌握性較好，進而增強對台灣地區的預報能力。

為了更加強對台灣的預報能力，我們利用多變數模式 (OLR 與 850 hPa 的渦度場) 做改進，結果在較為明顯的特徵區與單一變數的預報驗證差異甚小。整體而言，多變數模式與單一變數模式表現相當。

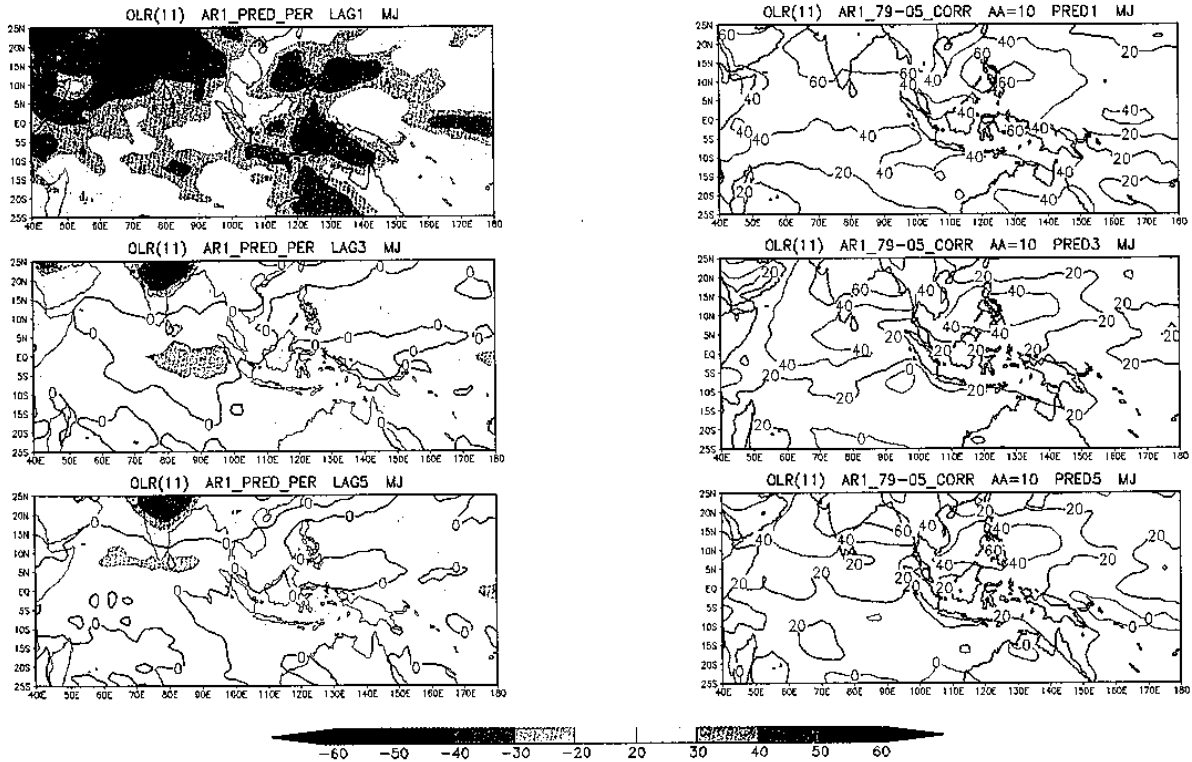
模式的改進對台灣地區的預報仍待改進；縮小範圍的結果，對模式的預報能力改善有限。既然多變數模式與單一變數模式的預報結果差異不大，我們會採用多變數模式與氣象局動力模式結合建置成一個季內尺度預報系統。

七、參考文獻

- 陳昭銘、陳仁會、呂芳川、郭敏冷、胡志文，2003：CWB GFS 模擬亞洲夏季季風環流之可預報度、準確度與海溫變化之關係。《大氣科學》，31，355-374。
- Barnston, A. G., M. H. Glantz, and Y. He, 1999: Predictive skill of statistical and dynamical climate models in SST forecasts during the 1997-98 El Niño episode and the 1998 La Niña onset. *Bull. Amer. Meteor. Soc.*, **80**, 217-243.
- Jones, C., Leila M. V. Carvalho, R. W. Higgins, D. E. Waliser and J.-K. E. Schemm. 2004: A Statistical Forecast Model of Tropical Intraseasonal Convective Anomalies. *J. Climate*, **17**, 2078-2095.
- Michaelsen, J., 1987: Cross-validation in statistical forecast models. *J. Climate. Appl. Meteor.*, **26**, 1589-1600.
- Mo, K. C., 2001: Adaptive filtering and prediction of intraseasonal oscillations. *Mon. Wea. Rev.*, **129**, 802-817.
- Saha S., S. Nadiga, C. Thiaw, J. Wang, W. Wang, Q. Zhang, H. M. Van den Dool, H.-L. Pan, S. Moorthi, D. Behringer, D. Stokes, M. Pena, S. Lord, G. White, W. Ebisuzaki, P. Peng, P. Xie, 2005: The NCEP Climate Forecast System. *Submitted to the J. Climate*.
- Waliser D. E., C. Jones, J.-K. E. Schemm and N. E. Graham, 1999: A Statistical Extended-Range Tropical Forecast Model Based on the Slow Evolution of the Madden-Julian Oscillation. *J. Climate*, **12**, 1918-1939.
- Yuval, 2000: Neural Network Training for Prediction of Climatological Time Series, Regularized by Minimization of the Generalized Cross-Validation Function. *Mon. Wea. Rev.*, **128**, 1456-1473.

持續性預報 (五六月)

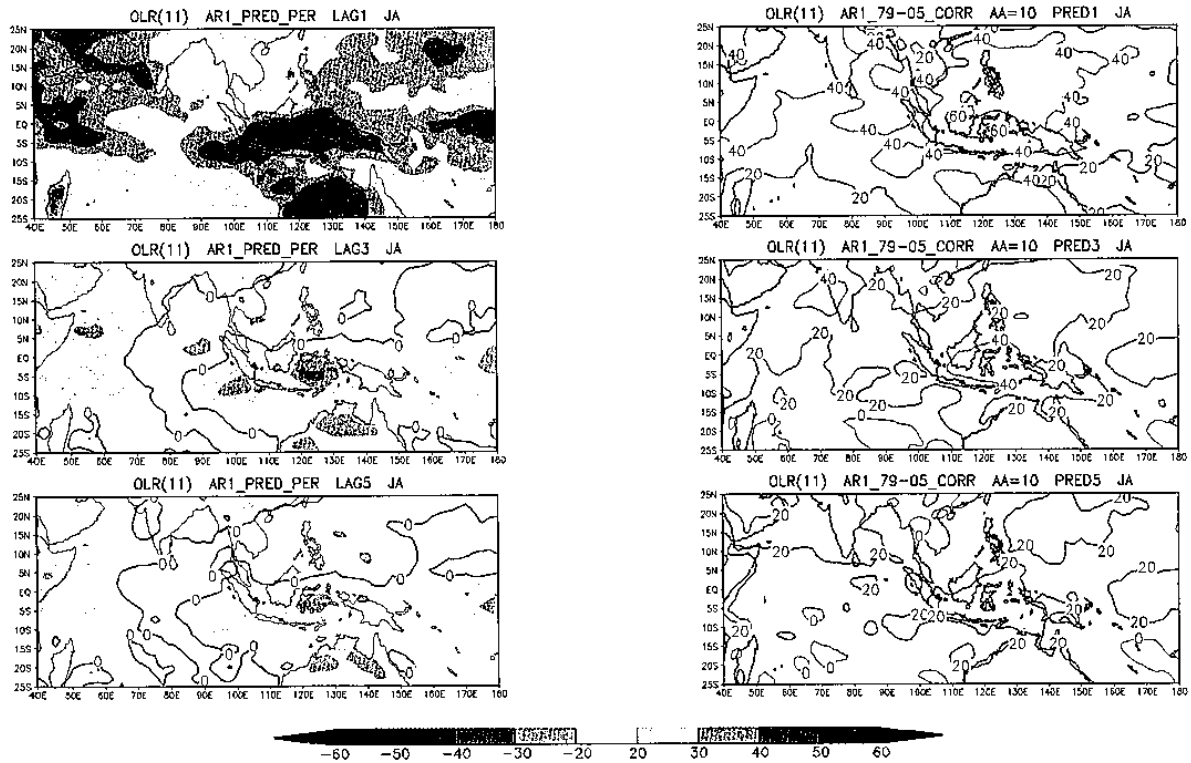
預報驗證 (五六月)



圖一：使用單一變數 OLR AR1(II)模式的預報結果比較。左右排分別是五六月的持續性預報及預報驗證，由上到下則分別為第 5、15、25 天預報。在持續性預報圖中，陰影表示 OLR 自身時間延遲的相關係數（單位為%）；但在預報驗證圖中，線條表示預報與觀測 OLR 的相關係數（單位為%），陰影表示通過 95%的可信度。

持續性預報 (七八月)

預報驗證 (七八月)



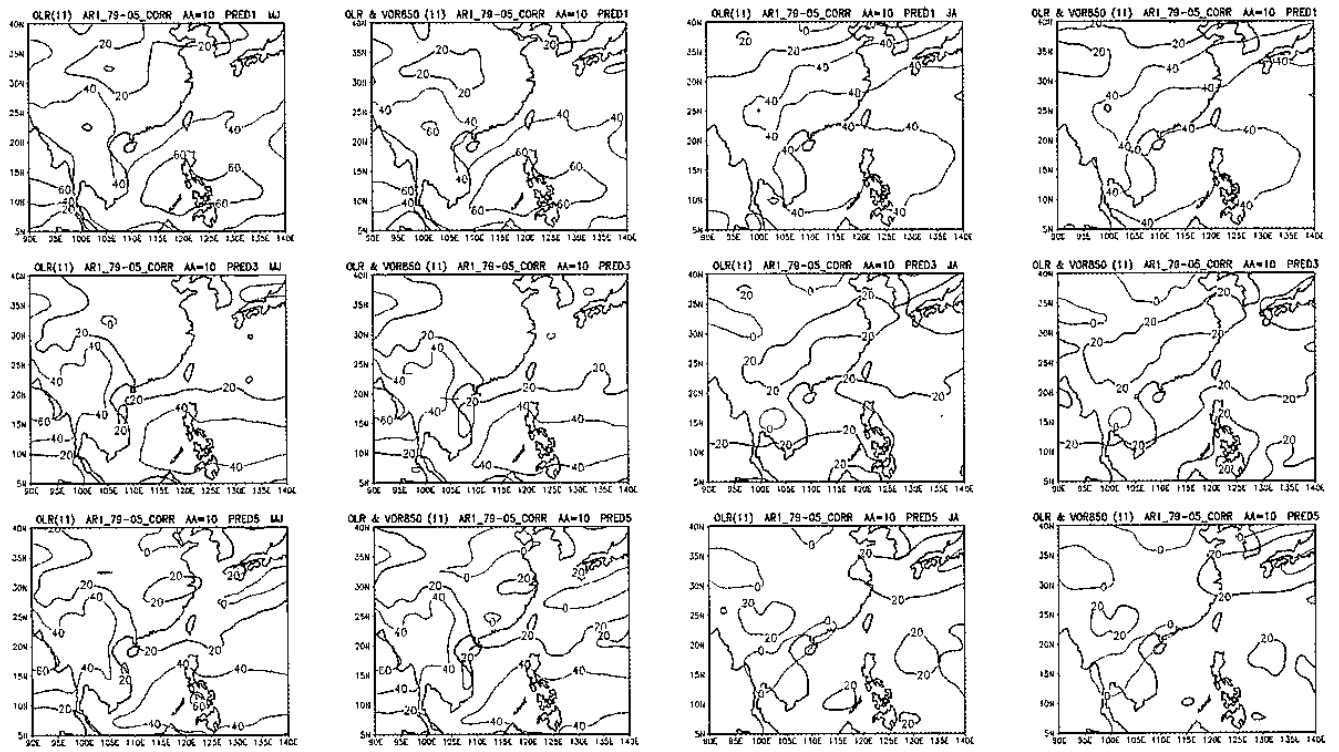
圖二：同圖一的解說，但是比較月份變成七八月。

單變數模式 (五六月)

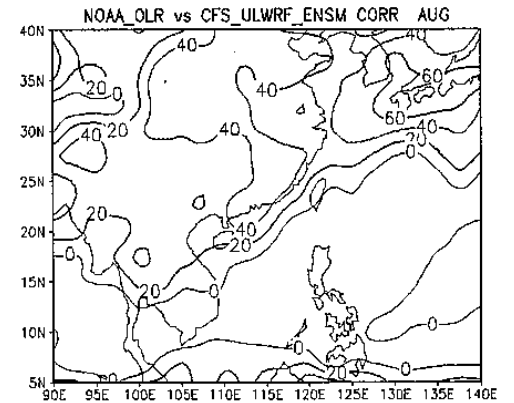
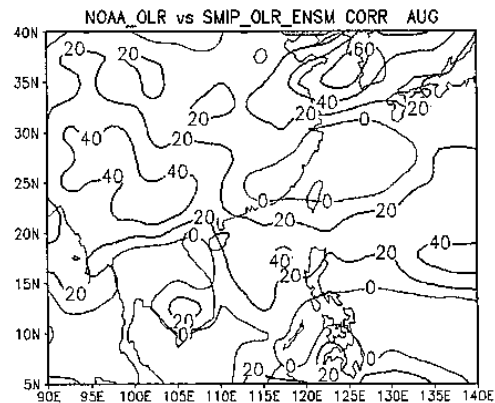
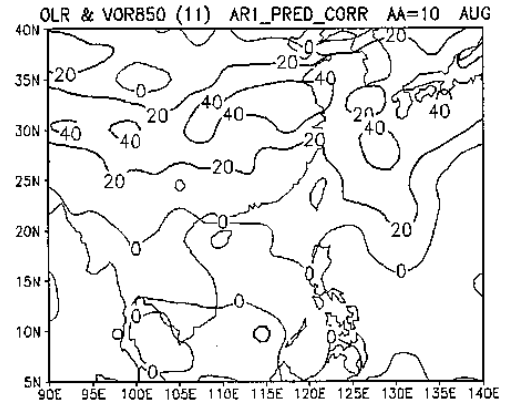
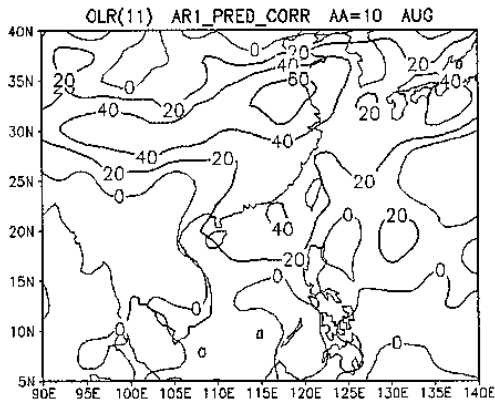
雙變數模式 (五六月)

單變數模式 (七八月)

雙變數模式 (七八月)



圖三：四欄皆為預報與觀測OLR的相關係數(以線條表示,單位為%),陰影表示通過95%的可信度。預報區域為90°E~140°E, 5°N~40°N。左二欄為五六月,右二欄為七八月;一、三欄為單變數模式的預報,二、四欄為雙變數模式的預報。



圖四：三個不同模式在八月份的預報與觀測OLR的相關係數(以線條表示,單位為%),陰影表示通過95%的可信度。左上：ARI(II) (單變數模式), 右上：ARI(II) (雙變數模式), 左下：中央氣象局SMIP, 右下：NCEP CFS。