

主成份分析應用在完全預報及數值模式預報 之迴歸統計之研究

陳建民 張智北
美國海軍研究學院
王惠民
中央氣象局科技中心

摘 要

天氣要素(如降水及最高最低溫)的客觀機率預報是數值天氣預報的一項重要產品,正常情形是利用數值預報之迴歸統計(Mode Output Statics;簡稱MOS)和完全預報法(Perfect Prog Method)來作,這兩種方法均包含了對過去數值模式預報場的迴歸分析。因為數值天氣預報分析及預報場均包含了大量資料,最理想而實際的作法是建立一套迴歸分析模式(Regression Model),而其中最重要的一個考慮就是如何適當的選取預報因子(predictors),其主要目的是當大部份的氣象資料都有高內相關(highly intercorrelated)時,把多重共線性(multicollinearity)所產生敏感的截斷誤差(round-off error)減至最小。因此,多重逐步的迴歸分析及類似的技術就必需建立在數值模式預報之迴歸統計模式或者是完全預報模式(perfect prog models)中。

本文將預報因子場(predictor fields)使用兩段式主成份分析(Principal Component Analysis;簡稱PCA)去除其多重共線性而保留其信賴度(reliability)。第一步用主成份分析處理每個預報因子場並選部份有用的主成份(principal component),選取的考量有二:1.將平均變異數中低於某一預設百分比之主成份捨去 2.然後求所得主成份與被預測要素(predictand)之相關係數(correlation coefficient),從係數中重新作評估,把先前捨去有相關意義的主成份再選進來,先前選取相關性低的主成份捨去。因為一組經適當處理過適合統計預報的氣象資料其間有意義部份與雜訊部份有其一有意義之比例,而所有被保留之主成份應接近於變異數截斷標準(variance cutoff criterion)。第二步中每個預報因子場所選之所有主成份都已標準化(normalized)且合併成一組資料,利用這組資料再作主成份分析而建構一個迴歸分析模式(regression model)。有一個潛藏的問題就是對於那些標準化過的末尾主成份(trailing components)的信賴度,這是一個有意義的問題。因為最大信賴度(maximization of reliability)與最大可預報度(maximization of predictability)間原本就是相衝突的。因此,我們另外作了變異數截斷(variance truncations)及相關性再評估(correlation re-evaluations)。因為所有主成份沒有內相關,所以上述程序中也就沒有多重共線性的問題。如此一來,用逐步迴歸分析以減少這種顧慮也就不需要了。但主成份分析的結果視預報因子場所取範圍大小而定,若有必要我們將作多次測試以找到每一個場的最佳範圍,而每一個案中,我們可以科學上及經驗上的知識作為尋找的指引,故尋找的次數理當較逐步迴歸分析少。

我們使用EC 1980-1989年 10,11,12,1 月的初始場建立一個迴歸分析模式作為台北冬天的降水完全預報。所選月份是根據王時鼎先生的『台灣氣候降水之形態及特徵』研究結果。資料有地面氣壓、850百帕U, V風場和溫度場、500百帕高度場及200百帕U, V風場,網格點解析度是 $2.5^{\circ} \times 2.5^{\circ}$,範圍 $10^{\circ} - 40^{\circ} N, 100^{\circ} - 140^{\circ} E$,每一個場有 221點;我們用此維度(dimension)來選取整組資料作主成份分析,這是為了方便程式維護及自由度評估(degree of freedom estimate),結果將直接在研討會中報告。

