

用全球海面溫度客觀預測台北月平均溫度

陳 圭 宏
氣象預報中心
中央氣象局

摘 要

使用NMC全球月平均海面溫度網格資料(40S-60N, 0E-180-2W, 2×2網格), 取4×4網格, 共1568個有效網格資料做為候選預報因子(predictors)。中央氣象局台北測站的月平均溫度為預報對象(predictant)。利用逐步迴歸(Screening regression)選取適當的預報因子, 建立複迴歸方程式, 做為預報方程。預測台北未來一個月的月平均溫度。

逐步迴歸選取預報因子的過程, 有三個很困難的基本問題: (一)何時要停止選取預報因子的動作, 是最佳時機? (二)所建立的複迴歸方程未必是最佳組合。(三)建立的複迴歸方程, 不一定能適用於新的資料組。文中用一些限制條件和人工挑選複迴歸方程, 以應付這些問題。

應用於實際的長期預報作業中, 從八十二年三月到八十三年九月, 計19個樣本可供校驗。如果把每年的月平均溫度以頻率分布分為偏冷、正常、偏暖三級, 則命中11次, 命中率57.9%, 技術得分30.3%, 絕對值誤差平均0.62°C。成績優於持續法和日本的複迴歸及類似法。

一、前言

赤道東太平洋地區, 海面溫度的異常高溫, 會導致中太平洋地區經向哈德雷環流(Hadley Circulation)增強, 緯向的Walker環流減弱, 而使得赤道東太平洋的地面氣壓下降, 赤道西太平洋的氣壓上升(Bjerknes, 1969), 此即為南方振盪(Southern Oscillation)。而赤道東太平洋的海面溫度異常, 是為艾尼紐(El Nino)現象。南方振盪和艾尼紐現象幾乎同時存在, 所以合稱為ENSO。由於赤道中太平洋海面溫度的上升, 對流活動增強, 對熱帶地區對流層有增溫作用, 整個熱帶地區的對流層厚度會增加(Horel and Wallace, 1981)。這種對流層厚度的增加, 以落後赤道東太平洋海面溫度正距平六個月相關最高(Pan and Oort, 1983)。Nitta and Yoshimura (1993)亦認為ENSO發生後約半年, 亞洲南部地區, 地面溫度的增暖有最好的相關。因為海面溫度和地面溫度有落後相關存在, 所以具有長期天氣預報的參考價值。

美國NMC使用MOS(Model Output Statistics)於客觀預測溫度、降水機率有很好的成績(Glahn and Lowry, 1972)。而MOS的發展主要是在建立複迴歸方程。所以本文將以全球海面溫度為項報因子(Predictors), 台北月平均溫度為預報對象(predictant), 建立複迴歸方程, 用於預報台北月平均溫度。

二、資料

全球月平均海面溫度資料是取自NMC, 1970-1990年, 21年的月平均海面溫度網格資料(40S-60N, 0E-180-2W, 2×2網格)。但因2°×2°網格的格點太多, 電腦無法負荷, 採取4°×4°網格, 共1568個網格點的海面溫度做為候選預報因子。台北月平均溫度取自中央氣象局台北測站的月平均溫度資料。

三、原理和方法

在建立複迴歸方程時, 首先取測站月平均溫度為預報對象, Y。網格點上的海面溫度為預報因子, X_i。則k元複迴歸方程可寫為

$$\hat{Y} = a_0 + a_1X_1 + a_2X_2 + \dots + a_kX_k \dots \dots \dots (1)$$

其中 \hat{Y} 是複迴歸方程的估計值, 和實際值Y有些微差異。所以要找尋適合的複迴歸方程, 必須使

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2 = \text{minimum} \dots \dots \dots (2)$$

其中數列 y_1, y_2, \dots, y_n 是Y的樣本, n =樣本數以本文中 y_i 即為台北每年的月平均溫度。或者是用還原變異數(Reduction Variance, R.V)可量度複迴歸方程的好壞(Glahn and Lowry, 1972)。

$$R.V = \frac{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 - \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}$$

$$= 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2} \dots\dots\dots(3)$$

其中 $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ Y 的平均值

由(3)，當複迴歸的估計值 \hat{y}_i ，趨近於實際值 y 時，R.V 趨近於 1。則複迴歸方程(1)幾乎可以完全解釋預報對象 Y。事實上，還原變異數 R.V 即是複迴歸方程的複相關係數 (multiple correlation coefficient)。基本上在方程式(1)中，當 k 值愈大，R.V 值會愈趨近於 1，可得到很好的複迴歸方程。但是各預報因子間，一般都有相關存在，會造成複迴歸方程式不穩定，不能應用於新的資料組。所以選用少數幾個互相獨立的預報因子於方程式(1)中，雖然沒有很趨近於 1 的 R.V 值，但反而較有實用性，可用於預報作業中。而要如何選取預報因子，可使預報因子間的相關性最小；又可達到最大的 R.V 值？本文是使用逐步迴歸 (Screening regression procedure) 選取預報因子的方法。

假設有 P 個候選預報因子 ($X_i, i=1,2,3 \dots P$) 可供選取，P 個預報因子中，選出和預報對象相關最好的預報因子 X_a 。接著繼續在 P-1 個預報因子中選取最適合的預報因子 X_b ，它可使複迴歸方程的 R.V 值增加最多。而在選取 X_b 時也就是在計算部分相關係數 (partial correlation coefficient) 最大的預報因子。接著又重複以上的步驟，在剩下 P-2 個預報因子中，再選取部分相關係數最大的預報因子 X_c ，進入複迴歸方程中。如此逐步選取較佳的預報因子，可得複迴歸方程(1)。但是 k 值要多大才剛好呢？k 值小，R.V 值不夠逼近 1，k 值大，又會使得複迴歸方程(1)變得不穩定，不能適用於新的資料組。所以必須在逐步選取的過程中，還要做複迴歸方程(1)的 F 檢定。直到 F 檢定失敗，逐步選取程序才停止。

以上是逐步迴歸選取較好預報因子的程序。但是仍然存在三個比較棘手的問題 (Klein, Lewis, and Enger, 1959)。

(一)雖然可用 F 檢定來終止逐步迴歸選取步驟，但是 F 檢定的有意義水平 (Significant level) 要多少才適當。太嚴格，複迴歸方程建立不起來。太鬆了，複迴歸方程又太不穩定。所以 F 檢定的有意義水平，仍很任意，由各使用者決定。

不過一般都要比 0.05 的有意義水平嚴格。

(二)雖然所建立的複迴歸方程是比較好的複迴歸方程。但仍不一定是最好的預報因子組合。例如：當 $P=100$ $k=3$ 就有 $C_3^{100} = 161,700$ 個排列組合，其他尚有 $k=2, k=4 \dots$ 等等，數目龐大的排列組合。所以最好的預報因子組合不一定能得到。

(三)雖然在逐步迴歸選取過程中，已經儘量在選入互相獨立的預報因子，但仍有可能是由於靠機會選入的，和預報對象間不一定有關係。所以尚不一定能應用於新的資料組，作為預報方程。

本文是使用 IMSL 的福傳 (FORTRAN) 副程式 (RSTEP) 逐步迴歸選取預報因子，建立複迴歸方程。為了應付 Klein 等所指的三個問題，本文用了一些限制條件，首先設定 $k=2$ ， $R.V > 0.65$ 且 F 檢定的有意義水平為 0.025。或者 $k=3$ ， $R.V > 0.75$ 且 F 檢定的有意義水平為 0.025。

因為全球海面溫度的網格點多達 1568 點，所以在建立好第一個複迴歸方程後，把相關最好的預報因子剔除，剩下 1567 個預報因子，仍可輕鬆的用逐步迴歸選取預報因子方法建立第二個複迴歸方程。如此再剔除相關次好的預報因子，可再建立第三個複迴歸方程。如此重覆五次。如果都通過限制條件，可得 5 個複迴歸方程。 $k=2$ 和 $k=3$ 都用同樣的步驟，希望從此儘可能得到最佳的複迴歸方程組合。而全球海面溫度取領先台北 2 個月到 12 個月，所以預測台北每個月月平均溫度最多可達 110 個複迴歸方程。但有 RV 值和 F 檢定的限制，預測台北一月的複迴歸方程只有 60 個，二月有 68 個，其他見(表一)。

複迴歸方程建立好以後，並不是每個都能用。必須有一次或兩次的預報成功經驗才適用。所以每個月又只剩下 10 個左右，甚至更少的預報值可用，而這些預報值即為台北的月平均溫度預測。

四、實際應用成效

因為完整的全球海面溫度資料是從 1970 年開始。而且又有最後的人工挑選出預測值。所以很難做大量的測試運算。不過從八十二年三月開始就已經使用此套模式於實際長期預報作業中。到今年 (八十三年) 九月，共有 19 個樣本可供校驗。圖一、為八十二年三月到八十三年九月的預測距平值 (虛線) 和觀測距平值 (實線) 的曲線圖，平均絕對值誤差為 0.62°C 。如果以頻率分布分為偏冷 (-)、正常 (0)、偏暖 (+) 三級，則由表二可看出共有 19 次的預測，命中 11 次，命中率達 57.9%，技術得分 30.3%。(戚等, 1978) 技術得分的算法如表三。比較對這段時間的預測，使用持續法，即這個月是高溫則下個月也預測高溫，這個月

是低溫預測下個月也是低溫的方法。則才命中7次，命中率36.8%，技術得分4.6%。再比較日本以500hPa高度場為預報因子，所建立的複迴歸方程或使用類似／反類似方法對日本各地區的預測，技術得分都很難達到10%（曾等，1994）（Ueno etc, 1990）。所以用全球海面溫度為預報因子，所建立的複迴歸方程組，對台北月平均溫度預測，具有長期預報的參考價值。

五、討 論

雖然在逐步迴歸選取預報因子過程中，可以輕易的挑出幾點海面溫度值和台北月平均溫度有很好的關係。可是這些海面溫度值是領先台北月平均溫度二個月到十二個月，而且分佈在各個角落，很難用合理的物理解釋兩者間的關係。也就是說所選出的複迴歸方程大部分仍是由於靠機會建立起來的，很難適用於新的資料組。尤其是越多元的複迴歸方程更是如此。所以本文限制最多為三元（ $k=3$ ）複迴歸方程。不像MOS或P.P（Perfect Prog.）所選入的預報因子都可以有很合理的物理解釋和預報對象的關係。所以這兩種方法選入的預報因子甚至可以多達10個，也就是十元複迴歸方程，而它們對於新的資料組仍很適用。本文中對於要挑出合適的複迴歸方程的策略是用人工挑出有兩年（1991, 1992）成功預報的複迴歸方程。認為有兩次的成功經驗，是由於機會產生的可能性會比較小，再去決定出較合理的預測值。但由人工挑選，又沒有一定的準則，只能用高相關和低誤差的原則去挑選而已。

雖然實驗快兩年，有較好的技術得分，但校驗的樣本只有19個，仍太少。而且圖一中，預測值距平（虛線）和觀測值距平（實線）間相關係數只有-0.01，可說是毫無相關，表示預測的數值仍無法掌握台北溫度的走勢，甚至會有出現大誤差的可能。所以每次做出的預測數值，沒辦法衡量正確的機會有多大。

六、結 論

使用NMC全球月平均海面溫度網格資料， 4×4 網格1568點，1970-1990，21年。利用逐步迴歸選取預報因子，可得多組複迴歸方程，預測台北月平均氣溫。實際應用於月長期預報作業中，到八十三年九月止，共有19個樣本，命中11次，命中率57.9%，技術得分30.3%。優於持續法和日本的複迴歸及類似／反類似法。不過因為沒有合理的解釋網格點上的海面溫度和台北月平均溫度的關係，複迴歸方程仍有很大的可能是由於靠機會產生的，也就是說所得的預測值會不穩定，不正確，也無法掌握住台北月平均溫度的走勢。

參考文獻：

- 曾振發，陳圭宏，林燕璋，1994：赴日本研習長期天氣預報技術。行政院所屬各機關因公出國人員出國報告書。
- 戚啓勳，嚴夢輝，1978：氣象統計學。復興書局。
- Bjerknes, J., 1969: Atmospheric teleconnections from the equatorial Pacific. *Mon. Wea. Rev.*, 97, 163-172.
- 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.
- Horel, J. D., and J. D. Wallace, 1981: Planetary-scale atmospheric phenomena associated with the Southern Oscillation. *Mon. Wea. Rev.*, 109, 813-829.
- IMSL, 1991: FORTRAN subroutines for statistical analysis.
- Klein, W. H., B. M. Lewis, and I. Enger, 1959: Objective prediction of five-day mean temperatures during winter. *J. Meteor.*, 16, 672-682.
- Klein, W. H., 1983: Objective specification of monthly mean surface temperature from mean 700 mb heights in winter. *Wea. Mon. Rev.*, 111, 674-691.
- Koizumi, ko., K. Kurihara, S. Maeda, 1992: Monthly mean temperature forecasting using an analog/anti-analog method. *Geophys. Mag.*, 44, 51-57.
- Nitta, T. and J. Yoshimura, 1993: Trends and interannual and interdecadal variations of global land surface air temperature. *J. Meteor. Soc. Japan.* 71, 367-375.
- Pan, Y. H. and A. H. Oort, 1983: Global climate variations connected with sea surface temperature anomalies in the eastern equatorial Pacific Ocean for the 1958-1973 period. *Mon. Wea. Rev.*, 111, 1244-1258.
- Ueno, T., K. Koizumi and K. Kurihara, 1990: Prediction of monthly mean temperature in Japan with multiple regression model. *Geophys. Mag.*, 43, 131-143.

Object forecast monthly mean temperature of Taipei
from global sea surface temperature

Guay-Hong Chen
Central Weather Bureau
Weather Forecasting Center

ABSTRACT

Use NMC global monthly mean sea surface temperature grid data set(40S-60E, 0E-180-2W, 2x2 grid mesh), picking 4x4 grid data. There are 1568 grids data totally, can be used as potential predictors. And the predictant is the monthly mean temperature of Taipei, Central Weather Bureau.

Using screening regression selection predictors method to build multiple regression equations which are used as a forecasting equation to predict next month mean temperature of Taipei.

There are three big problems in screening regression procedure:

- (1) What time the selecting procedure should be stop?
- (2) The builded multiple regression equation is not the best combination of predictors possibly.
- (3) The builded multiple regression equation may be not fit in the new data set.

In order to face these problems we give some restricted conditions and man selecting the suitable multiple regression equations.

Apply this model in the long range forecast operation from Mar. 1993 to Sep 1994. There are 19 cases can be verified. If we devide the monthly mean temperature frequence into cold; normal and warm three ranks. There are 11 correcting forecast cases, 57.9% correcting rate, and 30.3% skill score. The average absolute deviation is 0.62 C. The skill performace is better than persistence method; better than JMA's multiple regression and analog/anti-analog method.

表一 預測台北月平均氣溫，每個月的複迴歸方程組數

月	一	二	三	四	五	六	七	八	九	十	十一	十二
	60	68	82	84	74	89	109	109	78	87	103	76

表二 檢驗台北月平均氣溫預測的列聯表，(一)偏冷，(0)正常，(十)偏暖，橫列是觀測數，縱行是預測數。

		預測			
		—	0	+	
觀測	—	1	1	1	3
	0	0	6	3	9
	+	1	2	4	7
		2	9	8	19

表三 技術得分的算法和列聯表

		預測			
		—	0	+	
觀測	—	n_{11}	n_{12}	n_{13}	$n_{1.}$
	0	n_{21}	n_{22}	n_{23}	$n_{2.}$
	+	n_{31}	n_{32}	n_{33}	$n_{3.}$
		$n_{.1}$	$n_{.2}$	$n_{.3}$	n

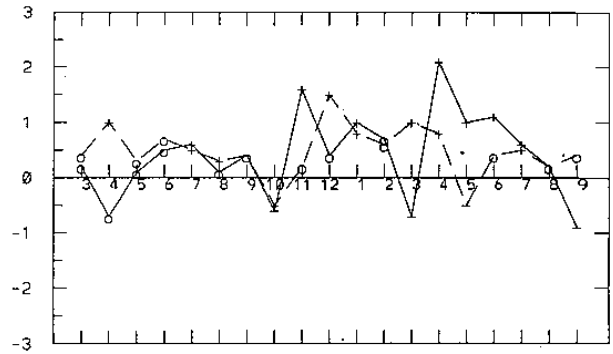
總數 $n = n_{11} + n_{12} + \dots + n_{33}$

命中數 $C = n_{11} + n_{22} + n_{33}$

命中率 $rate = C/n$

$$S = \frac{1}{n} (n_{.1} \times n_{1.} + n_{.2} \times n_{2.} + n_{.3} \times n_{3.})$$

技術得分 $Sc = [(C - S) / (n - S)] \times 100\%$



圖一 台北月平均氣溫的預測距平（虛線）和觀測值距平（實線），縱軸是距平值單位攝氏度，橫軸是月，3表三月，4表四月……等等。

