適用於推估臺灣日均溫度之統計降尺度方法

李柏宏
中央氣象局氣象預報中心

摘　要

統計降尺度方法中，以多元迴歸模型的運用最為廣泛，而建立多元迴歸模型的步驟中，最重要的部分則在於如何從大量的層場格點中建立解釋變數。本研究主要提出一個整合主成份分析和相關係數圖2種方法的優點，並且避免這2種方法使用上的缺點。

為了驗證本研究的可行性，文中探討模型推估4個氣象觀測站日溫度的狀況，此外也和4種建立解釋變數的方法進行比較。研究結果顯示推估臺北、臺中、高雄、花蓮測站之冬季和夏季日均溫度，均方根誤差平均都介於1至1.36度，而以高雄和花蓮測站的推估狀況最為穩定，受季節性的影響較小。推估值和日均溫度之間的相關程度分析顯示，冬季時大約都有88%的相關度，而夏季的相關度則較差，相關度大約在35%至50%之間。方法的比較上，冬季和夏季的技術得分平均都為正值，但冬季時的技術得分平均較夏季時的佳。

關鍵詞：統計降尺度、主成份分析、資料縮減、多元迴歸模型、溫度預測。

一、前言

為了了解天氣系統的變化和預測未來的天氣狀況，大氣科學方面的研究人員，根據流體力學、熱力學和大氣動力學等物理定律，建立出一些數學方程式，來描述大氣的複雜變化，並使用這些數學方程式，推估天氣的未來變化情況。此種推估的方法，在氣象上稱為數值天氣預測，而數值天氣預測會受到初始觀測場誤差、一部分大氣的物理過程和中小尺度天氣現象參數化的困難，及大氣層場解析度等因素影響，造成數值模式無法分析大氣層場格點(grids)與格點間所發生的物理過程，但小尺度的天氣變化資訊，在定量定點的天氣預測上也有著重大的影響，這也造成了降尺度方法的發展。降尺度方法的發展上，分成動力降尺度(dynamic downscaling)和統計降尺度(statistical downscaling)2種主要方法，而關於動力降尺度和統計降尺度2種方法的優缺點比較，在Fowler等(2007)有詳細的回顧，而本文主要針對統計降尺度方法的部分做討論。

統計降尺度方法是為了尋求感興趣的氣象觀測變數(如溫度、雨量、風速、雲量等)和數值大氣層場之間的統計相關性，建立合適的統計模型，以解釋網格點間的天氣變化，而感興趣的氣象觀測變數稱為反應變數(dependent variables)、數值層場的格點值稱為解釋變數(independent variables)。目前統計降尺度方法的研究，主要在建立大氣層場格點資料和氣象觀測站的觀測值，兩者之間的統計關係，而經常使用的統計方法有多元迴歸模型、正典相關分析(Canonical Correlation Analysis)、類神經網路(neural networks)方法(如Cavazos，2000、Hewitson，1997)，和類似法(如Karl等，1990、von Storch等，1993，和Murphy，1999)。這些統計降尺度方法中，以多元迴歸模型的運用最為廣泛，而建立多元迴歸模型的步驟中，最重要的部分則在於如何從大量的層場格點中建立解釋變數。經常被用來建立解釋變數的方法有3種，第1，利用某些指標，決定層場格點區域(如Zheng，2003)，並使用此區域的格點值當作解釋變數；第2，Walsh(1984)利用經驗正交函數(EOF)分析建立解釋變數；第3，計算層場格點和反應變數間的相關係數，建立相關係數圖(correlation map)，並選取統計檢定顯著的層場格點。對於這些層場格點，有些研究者會依據相關的背景知識，對不同的層場格點選取正相關的層場格點或是負相關的層場格點，並計算這些層場格點的平均值，當作解釋變數(如Chu等，2007、李，2009)。這些建立解釋變數的方法，都有一些使用上的缺點，例如，對於第1個方法，指標必須具有代表性，所決定的區域才有足夠的資訊，另外區域決定後，當層場格點資料相對於反應變數資料是相當龐大時，這會得到較差的迴歸模型參數估計值，並增加模型推估上的誤差；對於第2種方法，Huth(2002)發現在選取的大氣層場範圍較大時，格點所能提供的相關資訊，很可能會被大多不相關的格點資訊所稀釋；對於第3種方法，則必須要根據背景知識判斷，使用哪部分區域的正相關或負相關層場格點資料，或者正相關和負相關的區域都使用，而對於不同的研究人員，選取的區域不同，可能會得到不同的分析結果。另外，當決定好使用的區域後，僅僅利用區域平均值，是否可以代表這些格點的資訊，也是一個需要探討的問題。

因此本研究提出1個建立解釋變數的方法，並在多元迴歸模型的假設下，和上述所提到的方法進行比較，預期較好的建立解釋變數方法，將能降低多元迴歸模型的推估誤差。

本文的架構如下，第二章說明資料來源和研究方法、第三章為資料分析、第四章為結論。

二、資料來源與研究方法

(一) 資料來源

本文採用1981至2009年，氣象局4個局屬氣象觀測站(臺北、臺中、高雄、花蓮)之冬季和夏季日均溫度資料，而冬季定義為每年的1月、2月、和12月；夏季

圖1. NCEP重分析格點資料使用區域。

Fig. 1. The domain of the grid data used in this study (60°E~160°W, 0°N~60°N).

定義為每年的6月、7月、和8月。另外選定1981至2009年，在東亞區域NCEP重分析資料，共6個重要層場的大氣層場格點資料(2.5°x2.5°)，作為建立解釋變數的區域，此6個層場分別為海平面氣壓場、700百帕高度場、500百帕高度場、850百帕溫度場、850百帕垂直風場，和200百帕水平風場。如圖1，分析範圍為60°E~160°W、0°N~60°N。

(二) 主成份分析(principle components analysis)

主成份分析屬於多變量分析方法中的一種，其目標為經由線性組合的方法，得到相對較少且彼此互相獨立的主成份分數(principle component scores)，去解釋觀測資料中的大部分資訊，以降低大量資料的複雜度，而線性組合的權重則稱為主成份(principle components)，以下說明主成份分析方法過程。

1. 主成份分析統計理論

假設收集到的觀測資料矩陣D為，

$$D = \begin{bmatrix} x_{11}x_{12}\cdots x_{1n} \\ x_{21}x_{22}\cdots x_{2n} \\ \cdots\cdots\cdots\cdots \\ x_{p1}x_{p2}\cdots x_{pn} \end{bmatrix} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \cdot \\ \bar{x}_p \end{bmatrix},$$

n代表樣本個數、p代表變數個數。令p個變數為$X = (X_1 \cdots X_p)$，則變數期望值為，

$$\mu_X^T = (E(X_1), E(X_2), \cdots, E(X_p))、$$

變數之共變異數矩陣為，

$$\sum_X = E(X - \mu_X)(X - \mu_X)^T。$$

由於共變異數矩陣$\sum_X$為一對稱矩陣，故可對$\sum_X$進行對角化分解(diagonalization)，

得到

$$\sum_X = C\Lambda C^T$$

$$= \begin{bmatrix} c_{11}c_{12}\cdots c_{1p} \\ c_{21}c_{22}\cdots c_{2p} \\ \cdots\cdots\cdots\cdots \\ c_{p1}c_{p2}\cdots c_{pp} \end{bmatrix} \begin{bmatrix} \lambda_1\cdots 0 \\ 0\,\lambda_2\cdot 0 \\ \cdot\cdot\cdot\cdot\cdot \\ 0\cdots\lambda_p \end{bmatrix} \begin{bmatrix} c_{11}c_{12}\cdots c_{p1} \\ c_{21}c_{22}\cdots c_{p2} \\ \cdots\cdots\cdots\cdots \\ c_{1p}c_{2p}\cdots c_{pp} \end{bmatrix},$$

Λ為一對角矩陣，其對角線上之元素為$\sum_X$的特徵值(eigen values)，分別為$\lambda_1,\cdots\lambda_p$，$C$為一個正交矩陣，其第$j$行向量$\bar{c}'_j = (c_{1j},c_{2j},\cdots,c_{pj})$為對應特稱值$\lambda_j$的特徵向量(eigen vector)，$j = 1\cdots p$。

則主成份分數Z為Z = C'D，其中

$$Z = \begin{bmatrix} z_{11}z_{12}\cdots z_{1n} \\ z_{21}z_{22}\cdots z_{2n} \\ \cdots\cdots\cdots\cdots \\ z_{p1}z_{p2}\cdots z_{pn} \end{bmatrix} = \begin{bmatrix} \bar{z}_1 \\ \bar{z}_2 \\ \cdot \\ \bar{z}_p \end{bmatrix}、$$

$$\bar{z}_i = (z_{i1},z_{i2}\cdots z_{in})，i = 1\cdots p。$$

則$\bar{c}'_i$稱為第i個主成份、$\bar{z}_i$稱為第i個主成份分數。

因此主成份分數Z之共變異數矩陣為，

$$\sum_Z = \mathrm{cov}(Z) = \mathrm{cov}(C^TX) = C^T\sum_X C = \Lambda$$

而第j個主成份分數之變異數為

$$\mathrm{var}(\bar{z}_j) = \mathrm{var}(\bar{c}'_jX) = \bar{c}'_j\mathrm{var}(\bar{x}_j)\bar{c}_j$$

$$= \bar{c}'_j\bar{c}_j\mathrm{var}(\bar{x}_j) = \mathrm{var}(\bar{x}_j) = \lambda_j$$

則得到

$$\sum_{i=1}^{p}\mathrm{var}(\bar{z}_i) = \sum_{i=1}^{p}\mathrm{var}(\bar{x}_i)$$

因此藉由主成份轉換後，會得到p個互相獨立的主成份分數(無共線性)，並且p個主成份分數的變異數總和與觀測變數的變異數總和相等。另外，經由頻譜解構定理(Spectral decomposition theorem)，可以使得第1主成份分數保有最多的資料變異，次之為第2主成份分數，依此規則到第p主成份分數，即$\lambda_1 > \lambda_2 > \cdots > \lambda_p$，而主成份分數的變異數占變異數總和的比例，代表主成份保有資料的變異程度，稱為解釋度，以下本文都以解釋度稱之。因此在應用上，僅需要使用前幾個主成份分數就保有大部分的資料變異，並且達到資料縮減(data reduction)的目的，但對於要保留幾個主成份分數才足夠，以下提出2種廣泛使用的選取方法。

2.主成份個數的選取

主成份個數的選取並沒有一定的標準，目前有兩個廣泛使用的方法。

(1)累積解釋度

累積解釋度的計算方法:

$$R_K = \frac{\lambda_1+\cdots+\lambda_k}{\lambda_1+\cdots+\lambda_p}$$

當累積解釋度R_k大於一特定解釋度時，即選取k個主成份進行分析，此特定解釋度為研究者決定，並沒有一特定標準。

(2)陡坡圖(scree plot)

陡坡圖的繪製方法，橫軸為主成分的次序，即1到p個主成份；縱軸為相對應主成份的變異數。將圖上每個座標連接起來，從圖中找出曲線突然轉折彎曲(elbow)的地方，該轉折以上的特徵值所相對應的主成份，就是欲選取的主成份個數。

(三) 多元迴歸模型(multiple regression analysis)

多元迴歸分析是利用多個解釋變數去解釋和推估反應變數的統計分析方法，以下說明多元迴歸模型的假設、參數$\bar{\beta}$的估計方法。

假設反應變數為y、k個解釋變數為$w_1 \cdots w_k$，則多元迴歸模型為，$y_i = \beta_0 w_{i1} + \beta_1 w_{i2} + \cdots + \beta_k w_{ik} + \varepsilon_i$，$\varepsilon_i \overset{iid}{\sim} (0,\sigma^2)$，$i = 1 \cdots n$，其中$n$代表樣本個數，$\varepsilon_i$為一平均數為0、變異數為$\sigma^2$的白噪音(white noise)。將上列多元迴歸模型改寫成矩陣形式，則模型為，

$\bar{Y} = W\bar{\beta} + \bar{\varepsilon}$，其中

$$\bar{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}，W = \begin{bmatrix} 1 w_{11} w_{12} \cdots w_{1k} \\ 1 w_{21} w_{22} \cdots w_{2k} \\ \cdots\cdots\cdots\cdots\cdots \\ 1 w_{n1} w_{n2} \cdots w_{nk} \end{bmatrix},$$

$$\bar{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}，\bar{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}。$$

參數$\bar{\beta}$估計的方法，使用最小平方法(least squares method)進行估計，則最小平方法估計量$\hat{\bar{\beta}}$為，$\hat{\bar{\beta}} = (W^T W)^{-1} W^T \bar{Y}$。

使用收集到的資料，得到迴歸參數估計量$\hat{\bar{\beta}}$後，可利用新收集到的解釋變數$\bar{w}'_f = (w_{f1}, w_{f2}, \cdots, w_{fk})$，對未來的反應變數$y_f$進行推估，其推估值為$\hat{y}_f = \bar{w}'_f \hat{\bar{\beta}}$。

(四) 建立解釋變數的方法和推估流程

此節說明本研究擷取層場格點資料，建立解釋變數的方法，並利用這些解釋變數建立多元迴歸模型。首先說明建立解釋變數的方法，利用觀測站之觀測資料和層場格點資料分別進行相關係數分析，建立相關係數圖，並決定相關度門檻值(threshold)進行層場格點資料的選取，即選取相關係數圖中，相關係數絕對值大於或等於相關度門檻值的格點資料，也就是正相關和負相關的格點，都將提供建立模型的資訊，當選取完格點資料後，再利用主成份分析，將各層場內所選取的格點資料進行線性轉換，建立互相獨立的主成份分數，並決定累積解釋度，用以選取主成份的個數，則完成了建立解釋變數的流程。

在建立解釋變數的流程中，有2個參數必須決定，第1個參數為相關度門檻值、第2個參數為累積解釋度。本研究使用交叉驗證(cross validation)的方法，來進行相關度門檻值和累積解釋度的選取。交叉驗證的流程為，假設使用n年的溫度資料和m個層場格點資料，則建立n個交叉驗證集，每1個交叉驗證集內，又被分成進行校驗的測試資料(testing sets)，和進行模型建立的訓練資料(training sets)。交叉驗證集的建立方法為，第1個交叉驗證集由第1年溫度資料當作測試資料進行校驗，其餘n-1年當作訓練資料進行建模；第2個交叉驗證集由第2年溫度資料當作測試資料進行校驗，其餘n-1年當作訓練資料進行建模，依照此規則建立其餘n-2個交叉驗證集，在固定不同的相關度門檻值α和累積解釋度γ下，進行交叉驗證分析，使得平均校驗準則最佳之α、γ即是建議選取的參數值。經由交叉驗證方法，選取合適的α、γ參數值後，即可建立多元迴歸模型。

這裡α、γ參數進行交叉驗證的範圍為$\gamma \in (0, 0.05, \cdots 0.95)$、$\alpha \in (0, 0.05, \cdots, u)$，其中$u = 0.05 \times [\min(\{\rho_i | i = 1 \cdots m\})/0.05]$、$\rho_i$為觀測資料和第$i$個層場格點資料中最大絕對

值，[·]為高斯符號。校驗準則使用均方根誤差(root mean square error；RMSE)，均方根誤差計算方式為，假設觀測資料為y_t、模型推估值為$\hat{y}_t$，其中t=1…n，則，

$$\text{均方根誤差}:\sqrt{\frac{\sum_{t=1}^{n}(y_t-\hat{y}_t)^2}{n}}。$$

注意到儘管使用m個大氣層場資料，這裡僅使用一個所有層場共用的相關度門檻值(overall threshold)，有以下的考量。第1，當m增加時，經由交叉驗證挑選各別層場的相關度門檻，是件耗費時間的事，並且不保證能降低RMSE；第2，高相關性的層場格點資料，當α值增加時，被挑選到的格點資料，會比低相關性的層場格點資料多，以減少相關性較低的格點資料，進入建立模型的過程。以上描述的建立解釋變數和多元迴歸模型過程，整理為圖2。

3.資料分析

此章分析本研究所提出之方法，並進行3個方向的分析。第1，交叉驗證分析的結果探討；第2，本文建立解釋變數的方法在推估日溫度上的結果；第3，和4種建立解釋變數的方法進行比較，觀察技術得分的結果。使用1981至2009年，臺北、臺

圖2. 建立解釋變數和迴歸模型之流程圖。

Fig. 2. The flow chart of establishing predictors and regression model building.

中、高雄、花蓮觀測站之冬季和夏季日均溫度資料，和6個NCEP重分析層場格點資料，進行資料分析。本文利用1981至2000年的資料進行多元迴歸模型的建立，並推估2001年的冬季和夏季日均溫度資料，再利用1981至2001年的資料進行多元迴歸模型的建立，並推估2002年的冬季和夏季日均溫度資料，以此方式持續推估到2009年。

(1)交叉驗證結果

經由交叉驗證分析後，可以繪製出在各個相關度門檻值和累積解釋度組合下的RMSE平均圖，再藉由此圖來判斷合適的α、γ參數值。圖3為臺北觀測站，1月和6月均溫交叉驗證結果，所繪製出的RMSE平均圖。圖3中可以發現2個特徵。第1，臺北站1月均溫和6個大氣層場格點資料之間的相關性較高，因為u值最大達0.5；而6月均溫和6個大氣層場格點資料相關性則較低，因為u值最大僅達到0.35，這顯示6個大氣層場中，至少有一個層場和6月均溫的相關性較低，而研究中發現僅200百帕水平風場和6月均溫的相關程度較低外，其餘大氣層場最高都有0.4至0.55之間的相關度；第2，在1月時，RMSE平均圖大約在累積解釋度0.6處，分成明顯的左右兩部分，左半部呈現出較大的RMSE平均，而右半部除了在低相關度門檻值和過高的累積解釋度時，呈現較大的RMSE平均外，其餘則是呈現出較小的RMSE平均。這顯示對於推估1月均溫，累積解釋度的影響比相關度門檻值還來的大，因為在任何相關度門檻值時，低累積解釋度的RMSE平均，並沒有比高累積解釋度的RMSE平均低，但是在低相關度門檻值時，選取過高的累積解釋度，可能會使得

圖3. 臺北測站1月和6月，交叉驗證結果。

Fig. 3. The cross validation result of Taipei station in January and June.

模型中包含過多的雜訊，造成較差的推估結果。6月時的RMSE平均圖，也可以觀察到圖中的RMSE平均，是由左到右逐漸減少，大約在累積解釋度0.5處，也分成左右2部分，同樣的，這也顯示出累積解釋度的影響比相關度門檻值還來的大，而圖中低相關度門檻值和過高累積解釋度的區域，則呈現出較大的RMSE平均，並且比1月來的大。

文中省略臺中、高雄、花蓮觀測站的交叉驗證分析圖，和臺北測站2月、12月、7月，和8月的交叉驗證分析圖，這些省略的交叉驗證分析圖，大致上和臺北觀測站的1月、6月交叉驗證分析類似，並且都有上述2點特徵，因此僅提及臺北觀測站的情況。歸納上述的探討，在建立解釋變數時，使用高相關度門檻值和高累積解釋度，預期會有較好的推估結果；而使用低相關度門檻值和過高的累積解釋度，則預期會使得模型包含過多的雜訊，而造成較差的推估結果。文中使用在交叉驗證下，使得RMSE平均最小的相關度門檻值和累積解釋度組合，做為建立多元迴歸模型的標準。

(2)模型推估結果

此小節呈現，以本文提出的建立解釋變數方法，建立多元迴歸模型進行日均溫度推估的結果。觀察表1，臺北測站冬季的RMSE平均比夏季的RMSE平均較低，大約差0.1度，而以6月的RMSE平均較高，達1.5度，其餘月份則介於1.1至1.36度之間；臺中測站和臺北測站的情況則相反，夏季的RMSE平均比冬季的的RMSE平均較低，大約相差0.15度，比臺北站稍高，而以2月的RMSE平均較高，達1.4度，最低的RMSE平均為8月的1.04度，其餘月份則是介於1.15至1.35之間；高雄測站的RMSE平均，不論在冬季或是夏季的變化都是相當穩定，各個月份的RMSE平均介在1度左右；花蓮測站則是冬季和夏季的RMSE平均沒有太大差異，但1月和6月的RMSE平均稍高，分別為1.15度和1.11度，而其餘月份則介於1度左右。分析推估值和日均溫度之間的相關程度顯

表1、4個氣象測站各月份RMSE平均

Table 1、The average of root mean square errors of four stations in summer and winter.

月 / 測站	1	2	12	6	7	8	冬季	夏季
臺北	1.34	1.25	1.14	1.5	1.25	1.36	1.24	1.37
臺中	1.33	1.4	1.23	1.28	1.14	1.04	1.32	1.15
高雄	1.11	1.12	1.09	1.12	1.03	0.98	1.11	1.04
花蓮	1.15	1.06	1.03	1.11	0.95	0.93	1.08	0.99

示，冬季時的推估值和日均溫度的相關度較高，對於4個測站都有大約0.88的相關係數值；而夏季時的相關度則較差，相關係數值大約在0.35至0.5之間。

多元迴歸模型輔以本文建立解釋變數方法，在冬季和夏季的各個月份上，RMSE平均都不超過1.5度，尤其以高雄和花蓮測站的RMSE平均變化最為穩定，RMSE平均介於1度左右。

(3)比較不同建立解釋變數的方法

在前言中已介紹過，幾種經常被使用建立解釋變數的方法，但為了避免相關經驗和背景知識上的差異，這裡並不考慮和主觀決定層場格點區域建立解釋變數的方法進行比較，而是和藉由相關係數圖選取統計檢定顯著的層場格點，此種建立解釋變數的方法進行比較。此外，利用經驗正交函數建立解釋變數的方法則不另做探討，因為當α參數為0時，本文提出的方法，則等同於使用經驗正交函數建立解釋變數的方法，在此章的第1小節中，已顯示出在α參數為0時，預期會造成較差的推估結果，故不再對此種方法進行比較。故此小節在使用多元迴歸模型下，比較這些方法對於推估日均溫度上的差異。

以下分別說明將進行比較的4種方法。第1種方法為，計算相關係數圖，選取通過統計檢定(95%信賴水準)的正相關層場格點資料，計算這些層場格點資料平均值，再利用逐步迴歸法(stepwise selection)建立多元迴歸模型。第2種方法為，選取通過統計檢定的負相關層場格點資料，計算平均值，再利用逐步迴歸法建立多元迴歸模型。第3種方法為，選取通過統計檢定的正相關及負相關層場格點資料，分別計算正相關層場格點資料和負相關層場格點資料平均值，再利用逐步迴歸法建立多元迴歸模型。第4種方法為，直接利用逐步迴歸法，選取層場格點資料建立多元迴歸模型。以上是4種建立解釋變數方法的說明。另外，為了方便閱讀，以下分別以方法1、2、3和4代表此4種方法。

為了比較不同的方法，對推估日均溫度結果的差異，這裡使用技術得分(skill scores)，當作衡量本文的方法和其餘4種方法的差異。技術得分的計算方式為，假設兩個不同模型的RMSE平均，分別為$RMSE1$和$RMSE2$，則技術得分的計算方法為，

$$\frac{RMSE1 - RMSE2}{RMSE1} \text{x}100\%。$$

表2顯示，除了在8月時，方法1與方法3呈現負技術得分外，其餘都呈現正的技術得分，尤其冬季的技術得分顯著較大，而夏季的技術得分就不如冬季的明顯。觀察方法1到方法3的技術得分結果，顯示出使用正相關的層場格點或負相關的層場格點，還是同時使用正相關和負相關的層場格點平均值，作為建立迴歸模型的解釋變數，對於模型的推估結果是有很大的差異，因為若是對推估結果沒有影響的話，所得到的技術得分，應該是不會有太大的差異。這表示對於使用方法1到方法3，其中1種方法進行研究的人員來說，預期會得到不同的分析結論，這是值得注

表2、和4種建立解釋變數的方式比較之平均技術得分。

Table 2、The average of skill scores with four methods of establishing predictors in summer and winter.

月 / 方法	1	2	12	6	7	8	冬季	夏季
1	34.18%	33.19%	34.26%	8.32%	6.12%	-0.76%	33.88%	4.56%
2	41.26%	48.48%	42.12%	4.27%	10.23%	2.66%	43.95%	5.72%
3	32.68%	33.72%	34.92%	5.87%	5.42%	-1.62%	33.78%	3.22%
4	95.28%	95.12%	95.54%	94.83%	93.6%	96.16%	95.31%	94.87%

意的。另外，可以注意到方法4的技術得分很高，而產生這個現象的原因，就如同前言中所提及，當層場格點資料相對於日均溫度資料是相當龐大時，會得到較差的迴歸模型參數估計值，而較差的參數估計值，會導致較差的推估結果。

四、結論

在統計降尺度方法上，很重要的一個部分是在於如何從大量的層場格點中建立解釋變數，而在前言介紹的幾種方法，在使用上似乎已讓研究者產生困惑。例如，要如何選取重要的層場區域、要選取相關度多高的格點資料、要如何將選取的重要格點資訊整合、目前使用的大氣層場是否具有解釋力等的疑問，而不同的研究者會有不同的建立方法，而不同的建立方法也可能會導致不同的研究結論。因此本研究提出整合主成份分析和相關係數圖，這2種建立解釋變數的優點，並且避免這2種方法使用上的缺點。

本文提出的方法，有以下幾個優點。第1，當考量的層場範圍、層場種類、層場個數和資料來源都相同時，不會受到研究者的相關經驗和背景知識上的差異，造成研究結果的不同；第2，經由相關係數圖篩選後，大多數線性相關度較低的格點資訊並不會被選取；第3，當使用的大氣層場個數較多時，相關性較低的大氣層場，提供的格點數也較少；第4，降低主成份分數包含的雜訊量；第5，經由主成份轉換，單一層場的主成份分數彼此互相獨立，不再需要考慮單一層場內，解釋變數可能產生的共線性問題。對於第5個優點，必須進行額外補充說明，經由主成份轉換後，單一層場的主成份分數彼此互相獨立，但並不保證層場間的主成份分數會互相獨立，而整個研究過程中，筆者並沒有發現層場間的主成份分數存在共線性此一問題，因為當解釋變數存在共線性時，計算矩陣W^TW之行列式值會近似於0，即$|W^TW|\approx 0$，但在各個月份的迴歸模型建立

過程中，都呈現較大的行列式值，明顯並不近似於0，因此在本文中，解釋變數間並不存在值得注意的共線性問題。

迴歸模型推估日均溫度的結果，顯示高雄和花蓮測站的推估狀況最為穩定，受季節性因素的影響較小，並且根均方差值都介於1度左右。另外，在迴歸模型的假設下和4種方法進行日均溫度推估能力的比較。技術得分的結果顯示，本研究方法對於提升迴歸模型推估冬季日溫度的能力有很大的幫助，但是在推估夏季日溫度上，僅有稍微的提升。筆者認為一部分的原因是由於層場格點資料中，對於中小尺度天氣系統的資訊較少，而夏季天氣主要受到這些中小尺度天氣系統的影響，例如午後雷陣雨和颱風，這些天氣系統對於溫度上來說都是重要的影響因素，而吳明進(2005)由一系列的研究中也發現，臺灣的區域氣候其實較其它地區的氣候複雜很多，冬季較夏季有較高的可預測度。另一部分的原因，筆者認為和選取的大氣層場種類有關，當選取的大氣層場中有相關度較低的大氣層場時，應該剔除此大氣層場，以免影響其它層場變數選取過多相關度過低的格點資料(可能為雜訊)，降低了模型的推估能力。在實際應用上，可以先經由觀察各個層場變數的相關係數圖，將相關度比較低的層場變數予以剔除。

整個研究結果顯示，筆者驗證了前言的推論，「較好的建立解釋變數方法，將能提升多元迴歸模型的推估能力」。建議使用此建立解釋變數的方法，進行相關的統計降尺度研究。

五、參考文獻

Cavazos, T., 2000: Using self-organizing maps to investigate extreme climate events: An application to wintertime precipitation in the Balkans. J. Climate, 13, 1718–1732.

Chu, P.S., X. Zhao, C.-T. Lee, and M.-M. Lu, 2007: Climate prediction of tropical cyclone activity in the vicinity of Taiwan using the multivariate least absolute deviation regression method. Terr. Atmos. Ocean. Sci., 18, 805-825.

Fowler, H.J., Blenkinsop, S. and Tebaldi, C. 2007. Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. International Journal of Climatology, 27(12), 1547-1578.

Huth, 2002. Statistical downscaling of daily temperature in central Europe. Journal of Climate. v15. 1731-1742.

Hewitson, B.,1997: GCM derived climate change impacts on regional climate variability. Preprints, Fifth Int. Conf. on Southern emisphere Meteorology and Oceanography, Pretoria, South Africa, Amer.Meteor. Soc., 24–26.

Karl, T. R., W.C. Wang, M. E. Schlesinger, R. W. Knight, and D.Portman, 1990: A method of relating general circulation model simulated climate to the observed local climate. Part I: Seasonal statistics. J. Climate, 3, 1053–1079.

Murphy, J., 1999: An evaluation of statistical and dynamical techniques for downscaling local climate. J. Climate, 12, 2256–2284.

von Storch, H., E. Zorita, and U. Cubasch, 1993: Downscaling of global climate change estimates to regional scales: An application to Iberian rainfall in wintertime. J. Climate, 6, 1161–1171.

Walsh, John E., 1984: Forecasts of Monthly 700 mb Height: Verification and Specification Experiments. Mon. Wea. Rev., 112, 2135–2147.

Zheng, X., and J. A. Renwick, 2003: A regression-based scheme for seasonal forecasting of New Zealand temperature. J. Climate, 16, 1843-1853.

吳明進，2005，台灣區域氣候模擬與預報研究-動力與統計降尺度，行政院國家科學委員會專題研究計畫，編號NSC93-2111-M-002-005。

李清縢，2009，西北太平洋颱風個數預報與檢驗，氣象學報，47，25-37。

A Downscaling Method Applied to Taiwan Daily Temperature Forecasting

Po-Hung Lee
Weather Forecast Center, Central Weather Bureau, Taiwan, R.O.C.

ABSTRACT

Among all the statistical downscaling methods, the multiple regression model is applied to weather forecast most extensively. How to use huge grids data is the most important issue in multiple regression model. In this study, correlation map and principle component analysis method are combined to create the predictors which contains the advantages of the two methods and avoids their shortcomings.

To verify whether the new method is able to apply, I compare it with four other methods for daily temperature forecasting of Taipei, Taichung, Kaohsiung, and Hualien stations in Taiwan in winter and summer. The result shows that the root mean square errors are all between 1 and 1.36, and which for Kaohsiung and Hualien stations are more stable, less affected by seasonal effects. Correlation coefficients for forecast values and daily temperatures are all about 88% in winter, 35% to 50% in summer. Skill scores are all positive in winter and summer, but the skill scores are better in winter.

Key words: statistical downscaling, principle components, data reduction, multiple regression model, forecast temperatures