

臺灣測站月均溫統計降尺度預報模型建置與初步評估

張凱鈞¹、陳苡甄¹、馮智勇¹、劉人鳳²、陳孟詩²
多采科技有限公司¹ 中央氣象局預報中心²

摘要

本研究使用氣象局局屬 TCWB2T2 與 NCEP CFSv2 兩組氣候模式的重預報(Reforecast)資料，並選擇 119°E 至 123°E、21°N 至 26°N 範圍內 1°×1°解析度的格點區域代表臺灣區域，針對局屬 25 個人工測站分別以空間內插、格點資料回歸、建構類比(Constructed Analogues，以下簡稱 CA)三種統計降尺度方法逐站、逐月建置月均溫預報模型，並以方均根誤差以及相關係數、標準化標準差初步評估各組統計模型的預報能力。

降尺度方法在不同月份以及不同測站的表現有所差異，經逐站/逐月建置預報模型並進行初步評估後發現，空間內插法平均而言表現較佳、空間回歸法誤差較小但資料離散度低、CA 法則能在前述兩方法表現不佳的 1 月有較大進步。考量以貝氏模型平均法(Bayesian Model Averaging，以下簡稱 BMA)進行氣候綜合預報，若引入 BMA 法計算的系集預報值離散度較大，則 BMA 法較能選出有預報能力者。本研究使用空間回歸法建置的模型的預報值離散度過低，並不適用於 BMA 法中，而空間內插以及 CA 法其預報值離散度接近觀測值，做為 BMA 法中校準模型(Calibration Model)的系集成員較為合適。

關鍵字: 測站預報、統計降尺度、建構類比法

一、前言

建置氣候預報與氣候變遷分析所需的統計降尺度模型常以格點為標的，例如：Wood 等人(2002, 2004)使用空間分解法(Spatial Disaggregation)、Maurer and Hidalgo(2008)將空間分解法與 CA 法相互比較、Abtzoğlu and Brown (2011)使用調整後的空間分解法與 CA 法進行日資料降尺度、Ahmed 等人(2013)則先進行空間分解後再校正模式偏差以保留模式本身變異度。上述研究皆以格點降尺度進行氣候分析，並非所有統計降尺度方法均可對空間分布不均勻的測站點進行降尺度。同一種統計降尺度方法對於日資料和月、季資料間的預報能力也不同。

臺灣地區進行氣候預報的難處在於：由於國土面積小，空間解析度較低的氣候模式上所能使用的格點數有限，完成格點資料的氣候預報後，針對更小區域的氣候預報需再使用空間內插或時間迴歸等統計方法建立模型，考量以上述流程建立之格點統計預報模型實為引入第 2 種統計降尺度方法，難以評估單一統計方法對於預報成果的影響。且對於測站降尺度而言，能有地表觀測值來直接驗證預報模型的表現，評估成果也較為可信。

105 年度「研發短期氣候綜合預報系統計畫(以下簡稱

短綜計畫)」成果為利用分位數應設法校正氣候模式月資料，本年度(106 年)短綜計畫則使用去年度經過偏差校正後模式資料，使用統計降尺度方法針對局屬 25 座人工測站建立預報模型並使用基本統計量大致評估其預報成效。

105 年度短綜計畫針對臺灣測站冬、夏季溫度預報，使用單回歸法以距離各測站最近格點之季均溫建立的校準模型、以及以氣候指標值和測站溫度建立的橋接模型(Bridging Model)兩者使用 BMA 法進行測站季均溫綜合預報，夏季的預報成果為 BMA 法在大多數測站選擇校準模型，但無論是校準模型或是橋接模型在夏季歷年預報值的離散程度皆不大，以致於夏季使用 BMA 法進行綜合預報成果並不明顯優於等權重平均預報成果。而校準模型使用的單回歸法，因過程中使用最小方差估計(Least Square Error Estimation)來確保預報值和測站觀測值有最小誤差，但同時也減少預報值時間序列的離散度，使得各年間預報值無明顯變化。本研究一方面使用回歸法以外的統計方法建模，觀察空間內插和 CA 法預報模型和回歸法的差異；一方面也調整回歸法的做法，使用測站鄰近的 4 個格點來建立預報模型，並評估三種方法所建置的校準模型是否適合做為 BMA 法的校準模型系集成員。

除了評估模式和觀測量值誤差的方均根誤差

(Root Mean Square Error, 以下簡稱 RMSE)、趨勢對應程度的距平相關係數 (Anomaly Correlation Coefficient, 以下簡稱 ACC) 外, 也將分析各預報模型的標準化標準差 (Normalized Standard Deviation, 以下簡稱 NSTD), 並以 NSTD 做為判斷預報模型能否成為以 BMA 法進行綜合預報之系集成員的依據。

二、資料來源

模式資料分別使用局屬第二代二步法氣候模式 (以下簡稱 TCWB2T2) 1982 年至 2011 年的 $1^\circ \times 1^\circ$ 之 120 組系集成員重預報資料, 以及美國國家環境預測中心 (National Centers for Environmental Prediction, 以下簡稱 NCEP) 產製的第二代氣候預報系統 (National Climate Forecast System Version 2, 以下簡稱 CFSv2) 的 1982 年至 2010 年之 $1^\circ \times 1^\circ$ 系集重預報資料。此 2 筆模式資料各自計算等權重平均後進行偏差校正與建立預報模型。本研究以分位數映射法 (Quantile Mapping) 校正模式偏差, 校正所使用的分析場為 NCEP 產製之氣候預測系統再分析資料 (Climate Forecast System Reanalysis, 以下簡稱 CFSR), 有將原始解析度 $0.5^\circ \times 0.5^\circ$ 調整為和模式解析度相符的 $1^\circ \times 1^\circ$ 。本研究選定 119°E 至 123°E 、 21°N 至 26°N 範圍內共 30 個格點的區域代表臺灣區域。

為比對測站統計降尺度模型的預報能力, 本研究使用「氣候變遷應用服務能力發展計畫」中使用克利金法補遺完成臺灣本島 25 座人工測站月均溫資料評估預報模型之預報能力。

三、研究方法與流程

使用空間內插、格點回歸與 CA 法三種方法建置預報模型。空間內插法基於林與盧 (2016) 的研究成果中可知: 以距離平方反比 (Inverse Distance Weight, IDW 法) 作為權重的空間內插法在 16 個測站所建立的統計預報模型是明顯優於利用預報因子回歸的 nSVD 法。在短綜計畫中除了 IDW 法, 也同時使用了雙線性內插法建立預報模型, 經過比對後發現 IDW 法與雙線性內插法建立的模型兩者誤差極小, 本研究將以雙線性內插法代表空間內插法與其餘兩種統計模型比較。

(一)、雙線性內插法: 假設 o 為欲內插測站點, 其座標為 (x_o, y_o) , o 座落在點 1~4 所圍成的四邊形範圍內, 如圖 1 所示, 點 1~4 的座標分別為 (x_1, y_1) 、

(x_2, y_2) 、 (x_3, y_3) 和 (x_4, y_4) 。假設 o 點的估計值為 v_o , 點 1~4 值分別為 v_1 、 v_2 、 v_3 和 v_4 , 則內插的步驟為:

首先由 (x_o, y_o) 作垂線, 求垂線 $x = x_o$ 與通過點 1 和 2 兩點直線: $y = y_1 + \frac{y_2 - y_1}{x_2 - x_1} \cdot (x - x_1)$ 交點 a , 以及與通過 3 和 4 兩點直線

$y = y_3 + \frac{y_4 - y_3}{x_4 - x_3} \cdot (x - x_3)$ 交點 b 座標, 得到:

$$x_a = x_b = x_o$$

$$y_a = y_1 + \left(\frac{y_2 - y_1}{x_2 - x_1} \right) \cdot (x_o - x_1)$$

$$y_b = y_3 + \left(\frac{y_4 - y_3}{x_4 - x_3} \right) \cdot (x_o - x_3)$$

計算 a 到 1 與 a 到 2, 以及 b 到 3 與 b 到 4 的距離,

$$d_{a1} = \sqrt{(x_a - x_1)^2 + (y_a - y_1)^2}$$

$$d_{a2} = \sqrt{(x_a - x_2)^2 + (y_a - y_2)^2}$$

$$d_{b3} = \sqrt{(x_b - x_3)^2 + (y_b - y_3)^2}$$

$$d_{b4} = \sqrt{(x_b - x_4)^2 + (y_b - y_4)^2}$$

可求得 a 、 b 兩點的內插值,

$$v_a = \frac{d_{a1} \cdot (v_2 - v_1)}{(d_{a1} + d_{a2})} + v_1 \quad v_b = \frac{d_{b3} \cdot (v_4 - v_3)}{(d_{b3} + d_{b4})} + v_3$$

計算 o 到 a 以及 o 到 b 的距離,

$$d_{oa} = |y_o - y_a|$$

然後可求得 o 點的內插值為:

$$v_o = \frac{v_a \cdot d_{ob} + v_b \cdot d_{oa}}{d_{oa} + d_{ob}}$$

判斷 o 點是否在四邊形 1234 內的邏輯是首先比較 x_o 與 x_1 、 x_2 、 x_3 和 x_4 值的大小, 其必須滿足 $x_1 \sim x_4$ 四值中至少有一個值大於 x_o , 且至少有一的值小於 x_o 的條件。比較 y_o 與 y_1 、 y_2 、 y_3 和 y_4 值的大小, 其同樣需滿足 $y_1 \sim y_4$ 四值中至少有一個值大於 y_o , 且至少有一

一個值小於 y_o 的條件。如果以上二條件均滿足, 則

可判定 o 點座落在四邊形 1234 中。在本研究中，雙線性內插法使用的資料為經過分位數映射法校正後的模式資料，並將資料本身扣去交叉驗證後的氣候值，轉為距平後才進行雙線性內插。

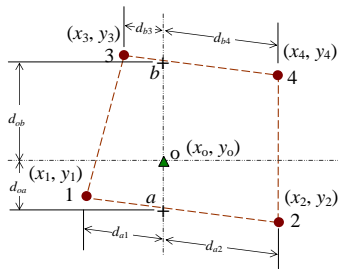


圖 1 雙線性內插法空間示意圖

(二)、格點回歸法：流程為找出距離各測站最近的四個格點，如圖 2 所示，並使用測站資料時間序列，將四個格點視為不同變數，兩者進行最小方差估計之複變數回歸，其求解矩陣運算式如下：

$$\begin{bmatrix} 1 & X_{yyr1}^1 & \dots & X_{yyr1}^4 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{yyrx}^1 & \dots & X_{yyrx}^4 \end{bmatrix} \times \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} = \begin{bmatrix} OBS_{yyr1} \\ \vdots \\ OBS_{yyrx} \end{bmatrix}$$

等號左邊的 X 矩陣為包含常數項之模式資料，上標 1 至 4 代表鄰近測站的 4 格點；下標 $yyr1$ 至 $yyrx$ 代表參與建模的年份， a_0 至 a_4 為欲求得的回歸係數， OBS 為觀測資料， OBS_{yyr1} 至 OBS_{yyrx} 為參與建模的觀測資料年份。求解過程為將模式對觀測值進行時間回歸建立模型。本研究中格點回歸除了找出鄰近測站的四個格點外，還額外使用了前進選擇法挑選四個格點中最有解釋能力的格點進行建模(短綜計畫中也有進行使用前進選擇法與否的測試，預報模型表現差異不大)，此外由於進行回歸即是偏差校正，因此在計算格點回歸時所使用的模式資料為未經分位數映射法偏差校正的資料，直接使用模式原值，不額外扣去氣候值計算距平。

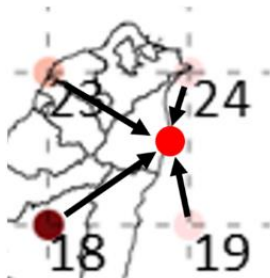


圖 2 蘇澳站以及用於建立回歸模型鄰近格點示意圖

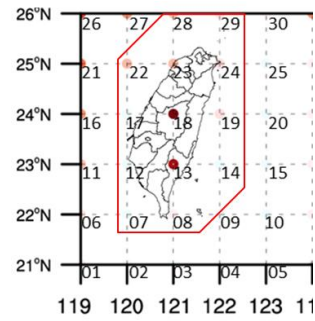


圖 3 涵蓋臺灣附近區域的 30 格點，與臺灣本島鄰近區域的 13 個格點(紅框)示意圖

(三)、CA 法：本研究使用的 CA 法為修改 Hidalgo 等人(2008)年提出的日資料降尺度方法，以符合月資料降尺度的需求。和日資料最大的差異在於篩選資料的方法由 RMSE 排名更改為使用前進選擇法挑選解釋能力較高者。此外 CA 法建立模型的格點範圍從一開始定義的 30 個格點經過初步測試後由於成效不佳，進一步挑選出格點範圍內有臺灣陸地的 13 個格點進行建模(如圖 3)。

月資料 CA 法其求解矩陣運算式如下：

$$\begin{bmatrix} X_{yr1,1} & \dots & X_{yrn,1} \\ X_{yr1,2} & \dots & X_{yrn,2} \\ \vdots & \ddots & \vdots \\ X_{yr1,30} & \dots & X_{yrn,30} \end{bmatrix} \times \begin{bmatrix} a_{yr1} \\ a_{yr2} \\ \vdots \\ a_{yrn} \end{bmatrix} = \begin{bmatrix} Y_{fct,1} \\ Y_{fct,2} \\ \vdots \\ Y_{fct,30} \end{bmatrix}$$

等號左邊的矩陣為透過前進選擇法選擇經偏差校正後，具有解釋能力的 n 年模式資料建模，縱向的 $X_{yrn,1}$ 至 $X_{yrn,30}$ 為具有解釋能力的第 n 年其臺灣區域 30 格點模式資料，橫向的 $X_{yr1,1}$ 至 $X_{yrn,1}$ 為具有解釋能力的第 1 年至第 n 年，右邊的 Y 矩陣則為模式預報年其臺灣區域 30 格點模式資料，中間的 a 矩陣則為欲求得的回歸係數。

在格點回歸及 CA 法的建模過程中，皆使用交叉驗證避免預報年的資料進入建置統計降尺度模型的流程中，例如若以 TCWB2T2 模式資料使用格點回歸在台北測站 1982 年 1 月預報月均溫值，則預報值是由 1983 年至 2011 年等其餘 29 年的模式 1 月均溫資料所估計出。

(四)、統計指標：本次的建置預報模型的成果將以逐站逐年的 RMSE 以及 ACC 來觀察預報模型對於測站資料趨勢的掌握及量值的差距，標準化標準差

(Normalized Standard Deviation, NSTD)用於判斷預報數據相對於測站資料的離散程度。其中 RMSE 若接近 0 則代表預報值和觀測值量值接近、ACC 若接近 1.0 為預報值和觀測值時間序列的趨勢對應相同、NSTD 若接近 1.0 則是預報值的離散程度和觀測值相符。將以上統計指標皆整理為如圖 3 的表格，能呈現單站逐月時間序列統計值的變化(沿 X 軸方向)，也能呈現單月份各站的表現(沿 Y 軸方向)。

四、研究成果與討論

圖 4、圖 5、圖 6 依序為空間內插、格點回歸、CA 法逐月份逐測站建立的統計降尺度月均溫預報模型，(a)、(b)分別為 TCWB2T2 以及 CFSv2 各自使用三種統計降尺度方法建立測站統計降尺度模型的結果，三張色階不同的圖表由上而下依序為 ACC、RMSE、NSTD。表 1 中的每一格數據為各預報模型各自的統計值將所有月份及測站加總平均，為該統計模型對於測站降尺度的整體表現。

由表 1 可知，不論空間內插、格點回歸、CA 法，CFSv2 相較於 TCWB2T2 而言，ACC 較高、RMSE 較低，顯示 CFSv2 相較於 TCWB2T2 而言較能掌握測站的溫度趨勢以及預報量值。空間內插法平均而言在三種降尺度方法中 ACC 最高(圖 4 的 ACC 圖明顯較圖 5 及圖 6 色階偏紅)，而其 NSTD 則因使用偏差校正後的資料而接近 1.0(分位數映射法的特色即在於使模式資料離散的程度接近 CFSR 分析場)，而使用預報年模式資料的空間內插法在 1、6、9、10 月的預報成果不明顯(如圖 4(a)(b))，顯示模式本身在這 4 個月的掌握度較差；空間回歸法雖因方法本身為最小平方方法，會讓 RMSE 為三者最小，但其 NSTD 在 TCWB2T2 中為 0.29、CFSv2 為 0.45，代表回歸法模型的預報值離散程度遠低於觀測資料，各年溫度的預報值差異甚小；CA 法雖然 RMSE 為三者最大，但 CA 法在建模時並不直接使用模式預報值建立模型，而是透過使用模式資料求出的回歸係數乘上測站觀測值進行預報，因而 NSTD 必定會接近 1.0，且 CA 法能夠在空間內插以及格點回歸於 1 月的預報能力不佳時有明顯改善，以 CFSv2 的 1 月模型最為明顯(見圖 6(b))。

本研究的 3 類統計降尺度預報模型中，空間內插與 CA 法資料離散程度接近觀測資料，格點回歸法的預報值離散程度則遠小於觀測資料。空間內插以及 CA 法所建置的模型依照先前的研究成果，將會因為

預報值離散程度接近而使 BMA 法較能個別給予不同權重；反之格點回歸法則因預報值離散程度遠小於觀測值，BMA 法的綜合預報成果和等權重預報的成果相比進步有限。本研究以短綜計畫日後的工作規畫而言，空間內插法以及 CA 法可用於建置臺灣測站統計降尺度月均溫預報模型。

參考文獻

1. 交通部中央氣象局, 2015: 氣候變遷應用服務能力發展計畫-104 年度氣候資料整集分析系統發展—觀測資料自動偵錯模組及輔助研判功能發展報告
2. 交通部中央氣象局, 2016: 研發短期氣候綜合預報系統計畫(1/2)案期末報告
3. 交通部中央氣象局, 2017: 研發短期氣候綜合預報系統計畫(2/2)案期中報告
4. 林昀靜、盧孟明, 2016: 氣象局月與季氣候預報系統統計降尺度預報技術評估, 105 年天氣分析與預報研討會
5. Abatzoglou J. T., Brown T. J., 2011: A comparison of statistical downscaling methods suited for wildfire applications. *International Journal of Climatology*. *International Journal of Climatology*
6. Ahmed K. F., Wang G., John S., Wilson A. M., Allen J.M., Horton R., Anyah R., 2013: Statistical downscaling and bias correction of climate model outputs for climate change impact assessment in the U.S. northeast: *Global and Planetary Change*
7. Hidalgo, H. G., Dettinger, M. D., Cayan D. R., 2008: Downscaling with constructed analogues: daily precipitation and temperature fields over the United States. *California Energy Commission, Sacramento, CA*
8. Maurer, E. P., Hidalgo H. G., 2008: Utility of daily vs. monthly large-scale climate data: an intercomparison of two statistical downscaling methods. *Hydrology and Earth System*

