

# 台灣自動氣象站氣溫資料補遺方法探討及網格化分析

陳雲蘭<sup>1</sup> 陳品妤<sup>1</sup> 詹智雄<sup>1</sup> 沈里音<sup>1</sup> 馮智勇<sup>2</sup> 劉家豪<sup>2</sup> 林佑蓉<sup>2</sup>

<sup>1</sup>中央氣象局

<sup>2</sup>多采科技有限公司

## 摘要

為了貼近實務應用所需的鄉鎮尺度空間解析氣象資訊，中央氣象局近年積極發展高解析網格化技術，藉以充分利用全台高密度自動測站量測數值來產製水平解析尺度可達 2.5 公里的細網格氣象資料。網格化結果可因所採用的網格化技術不同而有差異，但一般而言具一定水準的網格化技術多能表現資料的重要結構，同時亦能被隱含出個別技術的估計誤差特徵。本文將特別強調網格化過程引用參考觀測站資料品質及測站數量多寡對網格化結果的影響，由於自動站因遠地儀器維修不易及電訊傳遞有時不良等資料接收問題，其資料品質相對於人工測站較不穩定，特別在部分測站缺遺程度較高的情況下，非完整的觀測資訊將對網格化內插過程引入錯誤的空間統計關係。為避免資料樣本的不一致性，我們建議在網格化程序之前需對資料進行完整補遺或其他均一化的概念，本文將提出結合簡易插補法與利用鄰近站空間統計關係插補法的優點，分別處理非連續性缺值與連續性多筆缺遺值的問題，同時亦將引入數值預報模式資訊，利用 MOS 降尺度工具推估缺遺值，以協助討論補遺方法的不確定性。

## 一、前言

過去對於台灣氣候的了解，我們往往借助於二十幾個長期氣候人工觀測站資料，並可有效掌握大範圍綜觀氣候變化的特徵。不過台灣地形豐富複雜，海島氣候氣象萬千，不只受氣象綜觀系統影響，局地亦可能因地形條件而發展出與其他地點的特殊差異性，例如，因海陸風的影響，氣候特徵在沿海與靠山的地點可能有些不同，因地勢的影響，丘陵與平地、山前與山後對氣象綜觀系統的反應也可能不一。可想見僅靠二十幾個且多數又為臨海的人工站並不足夠描繪出台灣的小區域局地氣候特性。近年因環境與氣候變遷問題突現，各種民生事務所受環境因素影響程度受到關注，舉凡農作漁獲多寡、環境衛生品質、疾病傳染風險、水電資源調度、商品經濟強弱等等，各領域的分析皆愈來愈貼近小至鄉鎮空間尺度的民生應用需求，研究者對局地氣象資訊的需求日增，因為如果沒有能夠代表關注區域的環境資料，也就不能達成對關心問題的正確探索。對於台灣由縣市尺度邁入城鄉尺度的氣象分析，需要有更為廣布的高密度觀測資料。

拜自動化監測儀器科技發展之賜，中央氣象局在 1980 年代開始對全台佈建自動氣象站，從大台南及大台北開始，從西部到東部，由疏至密，過去卅年經過數期佈建及汰換計劃，目前已有超過 300 多個自動氣象站。其中最早的記錄資料是 1987 年開始，累積至今達 25 年以上。雖然現有測站有近半比例是近 10 年內所建置，但也有百餘站累積資料已超過

十數年，這些自動站的資料使我們想進一步了解城鄉尺度氣候特徵及其變化的希望變成可能。

為了滿足實務應用所需的鄉鎮尺度空間解析氣象資訊，中央氣象局近年積極發展高解析網格化分析技術，藉以充分利用全台高密度自動測站量測數值來產製水平解析尺度可達 2.5 公里的細網格氣象資料。其中，在民國 99 年起執行的「發展鄉鎮逐時天氣預報系統計畫」以克利金方法發展出一組細網格分析技術，目前已正式提供鄉鎮天氣預報作業使用，同時仍在相關計畫下持續進行細網格分析技術改進研究。本文工作進行臺灣氣候資料整合分析，是本局為提昇氣候與氣候變遷應用服務能力所規劃四年(103~106 年)工作計畫的一個子項部分，工作目標是透過對臺灣長期氣候資料的整合、處理及分析，以及產製長時間高解析度的氣候資料，逐步建立本局氣候資訊應用服務的基礎，進而提昇本局對臺灣氣候資料及氣候與氣候變遷資訊服務的能力與品質。今年(103 年)為此計畫開展首年，我們以溫度類氣候資料為優先，進行長期檔案資料的再檢查及缺遺問題處理，之後再以克利金方法產製城鄉尺度應用所需的高解析網格化資料。

網格化結果可因所採用的網格化技術不同而有差異，但一般而言具一定水準的網格化技術多能表現資料的重要結構，同時亦能被隱含出個別技術的估計誤差特徵。本文將特別強調網格化過程引用參考觀測站資料品質及測站數量多寡對網格化結果的影響，由於自動站因遠地儀器維修不易及電訊傳遞有時不良等資料接收問題，其資料品質相對於人工測

站較不穩定，特別在部分測站缺遺程度較高的情況下，非完整的觀測資訊將對網格化內插過程引入錯誤的空間統計關係。為避免資料樣本的不一致性，我們建議在網格化程序之前需對資料進行完整補遺或其他均一化的概念，本文將提出結合簡易插補法與利用鄰近站空間統計關係插補法的優點，分別處理非連續性缺值與連續性多筆缺遺值的問題，同時亦將引入數值預報模式資訊，利用 MOS 降尺度工具(陳等,2012)推估缺遺值，以協助討論補遺方法的不確定性。

## 二、資料及研究方法

本文說明我們先從氣溫單一變數開始整理臺灣長期氣候資料的工作。針對應用需求為提供長期氣候分析的鄉鎮尺度資料，我們的工作目標是使用具可靠度的客觀分析方法完成高解析網格化資料的產製。有關網格化所需的內插工具，我們採用先前計畫使用克利金方法所發展的網格化模組。不過，網格化的推估值乃取決於有觀測值的參考點，換言之，觀測值的品質及疏密程度決定了網格化的可信程度。因此我們在進行網格化之前，特別加強對觀測資料的檢查與處理，為了使每個測站能對區域氣候提供正確代表性，我們要求各站資料在時間的採樣頻率需為一致，例如皆為以 24 個小時代表一天的平均值，所有資料皆沒有缺遺。此外，參考測站在空間分布上也需夠均勻分布於全島。在此方向上，我們的分析乃以小時資料做為基礎，在分析年限方面，因本局自動氣象站在 1998 年佈建擴展至東部後才得以有較完整的全島分布，因此選定以 1998 年為分析起始年，如此，可完成至少 15 年度(1998~2012)的氣候資料整理，我們由部分應用端的回應了解這 15 年的長期資料已可滿足應用需求。以下分段繼續說明本工作選取網格化參考站的方法、對於經歷遷站撤站觀測資料的處理、對缺遺值的補遺方法，並再簡述所使用克利金網格化方法的特點。

### (一) 參考測站選取方法

自動儀器觀測資料難免缺失，但參考站觀測值的品質是網格化可信度的重要依賴，因此，除了挑選具長期觀測記錄的測站之外，我們亦要求其有效記錄值需達一定比例，以降低後續因必要補遺而引入誤差的影響程度。我們標定 1998 年至 2012 為分析範圍，以有效資料量達 85% 以上來挑選測站，最後可選出本局自動氣象站共 89 站，同時再加入資料較完整的本局人工測站 22 個，總計選出後續網格化代表測站共有 111 站。所挑選出的參考測站分布圖如圖 1。其中，1000 公尺以上的參考站共有 16 站，這些山區資料，將提供網格化時地景影響推估的重要資訊。需要再特別說明的是，所選的

測站中有 6 個站(三重、苑里、梨山、宜梧、新城、玉里)曾改遷移至鄰近地點，新的遷移地點如果不能代表原來觀測地點的氣候特性，則需要進行資料均一性處理，另外有 3 個站(大尖山、新屋、台南)則有因撤站或資料中斷 2~3 年的情形。在網格化所需參考站空間分布採樣均一的考量下，我們乃對這些撤遷測站的新舊資料以客觀分析方法進行合宜的介接，進行方式如下段說明。

### (二) 撤遷站資料處理方法

對於遷站導致非均一資料的問題處理，新舊站的關係是否有差異是第一個需要辨識的工作，雖然辨識新舊站的關係以簡單圖示直觀或可窺知大概，本研究仍利用客觀方法給予量化分析。如果有明顯差異，則需要調整，如果差異在一定可接受的小範圍，則可有 2 種選擇，(1) 不調整，暫時不處理此誤差，直接介接，(2) 仍使用客觀方法依關係式調整，以確保整體一致性。在本研究中，我們選擇不管誤差大小，皆一律使用客觀方法予以調整。當然，對於資料的任何後處理都可能存在一些誤差，此時，對於方法的有效運用及驗證就變得很重要。以下依二類情況說明對撤遷測站資料的處理方式：

1. 有鄰近替代站(有機會反應局地特徵，有較好代表性)：我們可由新站資料藉由本研究所使用的內插工具推估出舊站位置觀測的可能數值。利用空間統計法來求算舊站地點估計值，此時非常靠近的鄰近替代站通常會反應出最大權重。在本文所需處理的 9 個撤遷測站中有 5 個站(三重、苑里、宜梧、新城、玉里)用此方法完成新舊站資料介接。
2. 沒有鄰近替代站：仍使用空間統計法來求算因撤站或資料中斷期間的估計值，但因為沒有位置接近的鄰近參考站，若此待處理測站具有非其他測站可解釋的局地氣候特徵，將影響推估正確性。因此在過程中另外輔以調整系統性偏差值，調整值的決定乃根據過去使用所有參考站對此缺遺站的估計平均偏誤分析結果(遮蔽測站估計法，見三、(一)說明)。本文所需處理的 9 個撤遷測站中，台南、梨山、大尖山、及新屋 4 個站用此方法完成資料中斷的問題處理。(圖 2)

### (三) 觀測缺失資料補遺方法

如前所述，自動儀器觀測資料難免缺遺，雖然本研究已按資料高齊備率原則挑選代表站，但為了滿足後續分析能有統計樣本均一的基礎資料，避免各站因缺遺時間的分布不同，而影響對局部特徵的代表性，我們仍仔細探討各種補遺

方法，務使能儘可能的在一定可容許的誤差範圍內，達成使資料量均一的目標。

對於不齊備資料的處理，過去研究大多以前後資料求取平均處理，亦有以氣候值填補的做法。本研究認為對於只有 1~2 筆的非連續性缺值，使用前後筆觀測數值插補的方式確實可行，但對於更長時間的連續性多筆缺值狀況，因單一站點來自時間連續性的可用資訊已缺失，則需利用其他的資訊來源，例如透過與鄰近站相關程度來插補，或是利用動力模式分析的大氣資料以統計降尺度方法進行地面氣象數據推估。為驗證想法，本研究設計以下 4 種補遺方法進行比較，各方法的設計及說明如下：

1. 補遺方法一 (本文簡稱 Fill\_m1)：直接以 15 年氣候平均值填補，以 73 個候作為時間樣本分類依據，屬於簡易插補做法。
2. 補遺方法二 (本文簡稱 Fill\_m2)：仍利用氣候平均值，但加入對於當日天氣冷暖距平的考量。此為改進 Fill\_m1 的做法，對於只有 1~2 筆的非連續性缺值，除了以氣候值填補，再加入前後 2 筆觀測值的距平值，亦屬於簡易插補做法。
3. 補遺方法三 (本文簡稱 Fill\_m3)：乃為空間統計關係的插補做法。我們認為使用克利金方法插補(UK)再輔以平均偏誤修正的 Fill\_m3 可以更好處理連續缺值的狀況。有關輔以平均偏誤修正，其中一個原因是我們目前所使用的 UK 空間統計關係只建立在地理座標資訊，而尚未包含像是地表利用等其他影響因子，有可能不能反應出部份測站的局地特性。例如在冬季或夜晚時間，如果缺失值出現在市郊(例如淡水站)，以大多為市區的測站資料來推估此郊區站，雖然仍可一定程度地掌握合理變化趨勢，但容易出現偏暖的高估值。面對這種偏估問題，我們在方法設計中另外先以遮蔽測站去來分別求出推估各站的平均偏估值，如此所得空間統計關係的補遺值經過除去偏估成效良好，可大幅改進對長時間連續缺值的問題，彌補 Fill\_m2 的不足。
4. 補遺方法四 (本文簡稱 Fill\_MOS)：Fill\_m3 除了利用鄰近站的資訊進行補遺，同時使用推估值與實測值的平均誤差來提高對局地特性的成功掌握度，另一種獲取局地特徵資訊的方式是利用 NWP 資料，利用統計降尺度技術與地面站資料透過迴歸模型建置，可以統計模型得出缺失值的推估值。不過此方式需要可靠且具逐小時的

NWP 輸出值，目前並無去完全滿足對此計畫十多年長期資料補遺的需求。本文將以近年資料為例，使用歐洲 EC 模式的 MOS 預報值來觀察其在補遺方面的可利用性，以供將來進一步工作參考。同時，此結果亦可協助對 Fill\_m3 比對，提供一獨立分析結果來協助對本計畫最後挑選補遺方法的可參考性的評估。

#### (四) 網格化方法

本研究使用的網格化工具乃由克利金方法發展而來。除了用於產製高解析網格化資料之外，此網格化方法亦用於前述資料補遺過程所需的空間統計內插工具。在本局先前計畫，已詳細進行過網格化工具的探討，並說明建議採用克利金方法的理由(李，2009)。以下對此方法再略做簡介。

克利金方法有嚴謹完整的統計理論支持，乃是客觀地以數學方法來找出所分析物理量在站點之間的空間關係式。而所謂客觀，即是根據所分析出影響物理量的函式結構，利用求取最小均方差和加入不偏估條件來求解出權重係數。與其他以距離反比權重及事先給定影響半徑的內插方法相比，克利金方法的內插權重係數乃由資料的分布特性決定，而非事先給定，因此更具客觀性。

克利金方法的操作大致可分為 3 個步驟：(1) 結構分析 (Structure Analysis)：可決定物理量的影響變數的選擇及其型式，例如： $T(x,y,z) = b_1X + b_2Y + b_3Z$ ，影響變數的顯著性需通過檢驗。(2) 參數估計 (Parameter identification)：包含一階動差與二階動差的估計，前者為平均趨勢函數，後者為半變異圖函數。(3) 根據所分析結構，在一組不偏估條件下，以最佳化方式求算權重係數。

有關克利金方法與其他方法的優劣比較可見李(2009)及馮等(2012)等相關報告。

### 三、結果分析

由上一節所提研究方法介紹，可知本研究在時空的補遺皆充分依賴克利金方法，本文將對克利金方法對於此次應用資料的實際評估情形，給予量化的分析說明。因為只有在我們選用的客觀分析方法能表現出效益的前提下，才能期待其應用於補遺或網格化的適當性。本節分 2 部份說明本研究重要議題結果分析：(1) 克利金空間統計法在氣溫資料網格化的可信度分析 (2) 測站氣溫資料補遺方法比較分析。

#### (一) 克利金空間統計法在氣溫資料網格化的可信度分析

對於客觀分析網格化(或內插)方法的不確定性或可信度評估,一般會採用遮蔽測站法,或又稱去一估計法(Take-one-out)。亦即若有  $n$  個參考站(在本研究下, $n=111$  站),去一估計法每次挑出一目標站,利用其餘  $n-1$  站來估計其數值,共重覆作  $n$  次估計,最後以此  $n$  組估計值對觀測值的正確模擬能力來量化說明方法的可信度。經逐月分析,各月皆顯示克利金方法所得估計值能與對應觀測值有高度相關,以 1 月份為例,相關值在絕大多數的測站可達 0.9 以上,其中又有高比例在 0.95 以上,表示此方法確實能掌握所欲推估測站氣溫的變化趨勢(圖 3、圖 4)。另外,從均方根誤差(RMSE)來量化誤差程度,各月皆顯示估計誤差皆明顯小於可代表自然變異的標準偏差,表示誤差幅度確實有一個相當可接受的容忍範圍。以 1 月份為例,在去除平均偏誤之前,克利金方法(UK)對推估小時觀測資料的 RMSE 在平地站大約為 0.5~1 度之間,山區大多數測站則不超過 2 度。當去除平均偏誤之後(UK\_BC),平地及山地站的 RMSE 皆可進一步縮小 2 至 5 成。以 UK\_BC 的 RMSE 相對於資料標準差(SD)的比值來作為可信度量化指標,1 月份資料顯示克利金方法對此 111 個參考站的誤差比值多在 0.5 以下(圖 5)。

## (二) 測站氣溫資料補遺方法比較分析

前述克利金方法的遮蔽測站推估成效說明了本計畫所使用網格化(或內插)工具的可信度,本節接續要回頭來說明本文所縮編的資料補遺工作,如二、(三)節所述 4 個補遺方法中,Fill\_m3 即是使用克利金方法,以測站之間的空間統計關係來對各小時缺失值提供推估補遺,也是我們在面對連續性缺值問題時認為最可以倚賴的補遺方法。

補遺的動機是為了統計採樣的均一性,但不管是簡單插補或充分利用有效資訊插補,總是屬於推估值而難免有誤差,尤其是在缺遺程度較高的情形下,將更增加補遺的不確定性。不過,與現行作業直接使用不足月(日)資料計算的月(日)統計量相比,有經過一些合理判斷補遺處理的資料將有機會更具代表性。在 4 個補遺方法中,Fill\_m1 及 Fill\_m2 為只使用單一測站資訊的簡單插補,對 Fill\_m2 而言,簡單利用多年氣候值加入缺值前後偏冷或偏暖的調整值,非常合適用於單一小時缺失值,也有機會對 2 筆少量缺值做合理推估,至於更長的連續多筆缺值,因為已缺乏足夠鄰近的距平參考,僅能像 Fill\_m1 一樣直接以氣候值填補,若在此連續缺值時段剛好碰上比較波動的天氣,將因天氣偏離氣候值而使 Fill\_m2 填補值失宜,這種情形可以由引入空間關係補遺法的 Fill\_m3 來做較好的處理,其缺遺值的推估乃是建立在補遺測站與所有測站的空間統計關係,我們認為 UK 法再經過系統性的均誤修正後的修補值(Fill\_m3)可大幅改進對長時間連續

缺值的問題。此外,我們又引入統計降尺度方法 Fill\_MOS 方法的結果來比對差異,其資訊來源建立於 NWP,獨立於 Fill\_m3 的估計方法,因此可協助對 Fill\_m3 合理性的判斷。

透過檢視個別站不同缺遺程度的補遺結果,我們初步驗證了 Fill\_m3 的可用性,以下以個案資料為例說明。圖 6 圖示本研究 4 種補遺方法的比較,其中黑色圓點代表實測資料(OBS),應為佈滿 24 個小時整點,若圖中某小時未畫出,表示為缺值。藍色圓點代表以 MOS 方法推估的補遺資料,目前為每 3 小時一筆。淺色虛線及實線分別為 Fill\_m1 與 Fill\_m2 推估值,紅色實線是本研究以克利金空間統計法對缺值資料的處理結果(TIO\_BC,亦即 Fill\_m3)。

由圖 6 上圖取自天母站的個案可見該日有 4 筆缺值分散在 11 時、13 時、19 時及 21 時,除了 Fill\_m1 之外,其他 3 個補遺方法的估計是接近的。Fill\_m1 的插補值明顯與前後觀測值不連續,表示並不可靠,這就是在遇上顯著天氣冷暖變化時,直接以氣候值填補的做法可能會出現的問題。再看經過距平修正的 Fill\_m2,或是由空間統計關係求算的 Fill\_m3 皆與 Fill\_MOS 相近,且與前後觀測值的接續亦無突兀,表示相當可信。而幾個獨立方法之間的相互支持,亦驗證這些方法對補遺值的合理估計能力。

圖 6 下圖則以取自四湖站的個案來說明連續多筆缺值的補遺狀況,圖中案例為整日缺值,因此 Fill\_m2 與 Fill\_m1 皆同補以氣候值,但是 Fill\_m3 由來自其他測站的資訊可推算當日白天氣溫可能比平日偏低,與氣候平均的差值在最熱的中午時段甚至達到 5 度之多,這個偏冷的推估值不只可由 Fill\_MOS 得到支持,查閱天氣報告相關記錄顯示當日有雲雨帶影響,亦可更確認 Fill\_m3 估計值的合理性。如此,藉由合理有效的補遺方法,我們有機會把一個可以提供某日偏冷資訊的資料補救回來,即使仍可能存在一些不確定性,但只要誤差幅度不影響我們對折關心天氣變動的正確掌握,這些補遺數據就有符合期待的可用價值。

## 四、結語

雖然我們希望引入更多的自動站資料來描述鄉鎮尺度細網格的氣候,但直接使用難免缺值的自動站資料,可能因參考站非均一或時間採樣非均一而影響網格化結果的正確性,也對後續應用分析帶來錯誤解讀風險。本文介紹本局在「氣候變遷應用服務能力發展計畫」下準備氣候資料網格化的一個進展中的工作,重點放在資料的插補工具及做法。網格化資料品質依賴於參考測站,如果能表現區域性特徵的參考站出現缺遺,將影響網格化結果的代表性。本計畫要求網格化觀測參考站資料需為齊備,使網格化氣候資料能在時空採樣

均一的條件下完成，也就是在進行空間統計關係及插補時，能使用所選完整的 111 個參考站資訊，在時間統計量方面，亦能使用齊備的 24 小時資料。本文以網格化可信度分析說明本計畫所使用的網格化(插補)工具的適用性，並以實例說明應用在資料補遺的可用性，由於是仍在進展中的工作，僅就目前分析成果呈現，完整論述仍待後續完善分析。

## 五、參考文獻

- 李天浩，2009：應用克利金法建立高解析度網格點氣象數據之研究。交通部中央氣象局委託研究計畫成果報告。
- 陳雲蘭、劉欣怡、馮智勇、薛宏宇、劉家豪，2012：高解析度統計降尺度預報方法研究。天氣分析與預報研討會論文集編，中央氣象局
- 馮智勇、劉家豪、陳雲蘭，2012：客觀分析法地面溫度案例分析。天氣分析與預報研討會論文集編，中央氣象局。

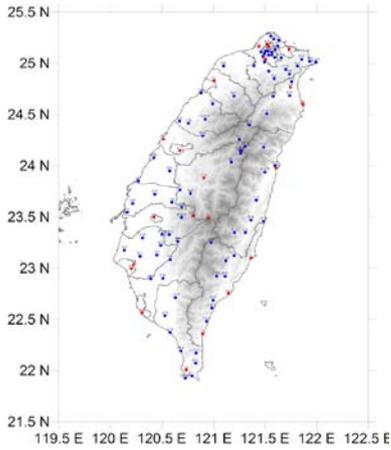


圖 1：本研究按資料高齊備率原則所挑選出的 111 個具備 15 年以上長期資料的網格化參考站。

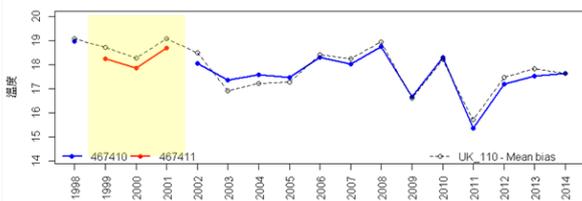


圖 2：以台南站 1 月份資料為例，圖示本研究以克利金空間統計法對介接氣候中斷資料的處理結果。其中藍色點或實線為台南站(467410)觀測資料，紅色實線為永康站(46742 / 467411)觀測資料，虛線則為以遮蔽測站法輔以調整平均偏差的推估值。陰影區為待處理資料介接的年份，我們將以虛線推估值進行中斷資料的修補。

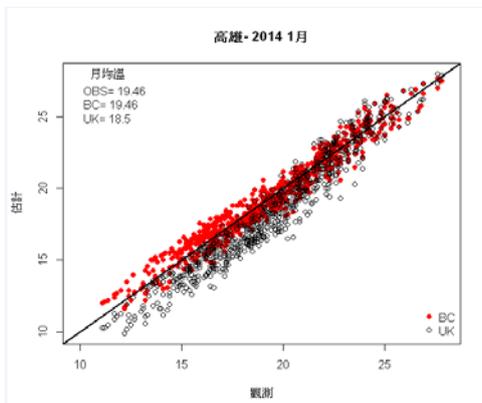


圖 3：以高雄站為例，圖示為透過遮蔽測站法說明克利金法對時間樣本的插補結果(UK，圖中空心圓點)，可見與實測值具高相關。再經調整系統性偏差的推估值(BC，圖中實心圓點)可以更進一步貼近實測值。

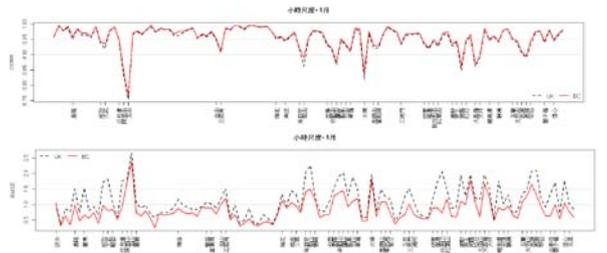


圖 4：圖示說明克利金空間統計法在氣溫資料網格化的可信度量化分析結果，使用遮蔽測站法分別對 111 個測站進行去一後推估，以 1 月份為例。上圖為各站推估值與實測值的相關係數值(COR)，顯示兩者具高度相關。下圖以均方根誤差(RMSE)來量化誤差程度。上下圖虛線皆代表未經調整的克利金(UK)推估成效，實線(BC)則代表經調整平均偏差的推估成效。

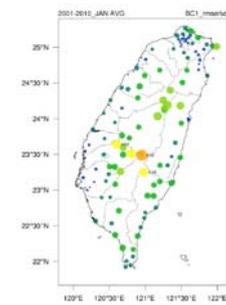


圖 5、以 UK\_BC 的 RMSE 相對於資料標準差(SD)的比值來作為可信度量化指標，1 月份資料顯示克利金方法對此 111 個參考站的誤差比值多在 0.5 以下

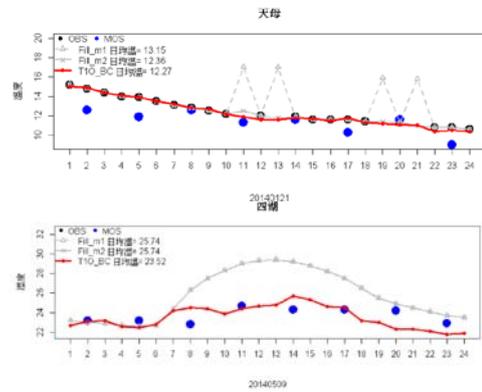


圖 6：以天母及四湖站為例，圖示說明本研究 4 種補遺方法的比較。其中黑色圓點代表實測資料(OBS)，應為佈滿 24 個小時整點，若圖中某小時未畫出，表示為缺值。藍色圓點代表以 MOS 方法推估的補遺資料，為每 3 小時一筆(2,5,8,...23)。淺色虛線為氣候值補遺方法(Fill\_m1) 推估值，淺色實線為在非連續性缺值情況下加上距平修正(Fill\_m2) 的推估值，紅色實線為本研究以克利金空間統計法對缺值資料的處理結果(TIO\_BC，亦即 Fill\_m3)。本研究認為使用克利金方法插補再輔以平均偏誤修正的修補值(Fill\_m3)合理，可以更好處理連續缺值的狀況，彌補 Fill\_m2 所不能處理的問題。