

# 高密度氣溫與雨量站觀測資料齊整率與合理性查核分析

呂致穎<sup>1</sup> 林淑卿<sup>2</sup> 陳品妤<sup>1</sup> 楊啟瑞<sup>2</sup> 沈里音<sup>1</sup>  
中央氣象局氣象科技研究中心<sup>1</sup> 中央氣象局第二組<sup>2</sup>

## 摘要

本研究為「臺灣長期氣候資料整集分析計畫」中的技術發展工作項目之一，此計畫工作已分別將氣溫和雨量資料整理累積近20年的自動氣象站及自動雨量站觀測值，對於氣溫長期資料缺遺問題，計畫首年（103年）已發展了合理補整方法；雨量長期資料特殊碼問題，計畫第二年（104年）已發展了合理處理方法，使城鄉尺度應用所需的高解析網格化資料可在參考測站時空採樣均一的條件下產製，計畫後續工作對氣溫和雨量資料的齊整率分析及資料品質檢測技術加以探討。

於本研究中，發現早期的資料齊整率較低，而近年來中央氣象局第二組（以下簡稱二組）對於自動測報系統觀測儀器的汰換及新增計畫，大幅提高了資料齊整率；另一方面，對於先前因通訊傳輸問題造成低齊整率的測站，二組亦透過進行資料補整作業程序，將正確可用資料填補回去，改善了整體資料齊整率。氣溫資料合理性查核分析分別導入了氣候離群值（Climate Outlier, CO）及迴歸離群值（Regression Outlier, RO）等2種方法來協助檢測，配合比對天氣系統，判斷是否為合理值。氣候離群值結果顯示於近6年的551筆可疑資料中，大部份為特殊天氣系統的合理值，僅有3筆為錯誤值；迴歸離群值結果顯示於近6年的17筆高度可疑資料中，有5筆為特殊天氣系統影響下的合理資料，另有12筆為觀測儀器穩定性不足造成的錯誤資料。雨量資料合理性查核分析則採用超越機率值（Exceedance Probability, EP）方法來協助檢測，藉由比對六個不同降水延時（Duration of rainfall）下，EP值為1%、5%的雨量閾值（threshold），作為判斷氣候極端降雨事件的標準，並配合比對天氣系統，判斷是否為合理值。結果顯示該方法確實可合理偵測極端降雨事件，亦有找出多筆錯誤雨量值的實例。透過以上3種合理性查核方法與天氣系統的分析，可累積對於可疑資料合理性判斷的經驗，以利後續資料品管流程的改進；並可將偵測出的錯誤資料回報給即時作業單位，以增進觀測資料的正確性。

關鍵字：資料齊整率、氣候離群值、迴歸離群值、超越機率值

## 一、前言

本研究工作進行臺灣長期氣候資料整集分析，屬於中央氣象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集、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。關於自動氣象站及自動雨量站在近期的資料齊整率與早期建立時相比，則有明顯的改善情形，究其原因主要為中央氣象局第二組（以下簡稱二組）對於自動觀測儀器的汰換及新增作業，大幅提高了資料齊整率；另一方面，對於先前因通訊傳輸問題造成低齊整率的測站，二組亦透過進行資料補整作業程序，將正確可用資料填補回去，改善了整體資料齊整率。計畫首年（103年）對

於氣溫長期資料缺遺問題，已發展了合理補整方法。計畫第二年（104年）對於雨量長期資料特殊碼問題，陳等（2015）已發展了合理處理方法；另針對氣溫長期資料的合理性查核分析，陳等（2015）提出透過觀測值與估計參考值的迴歸關係產出的迴歸離群值（Regression outlier）來協助資料正確性檢測，同時引入刪除式殘差（Studentized Residuals）的應用來量化觀測值的離群異常程度。使城鄉尺度應用所需的高解析網格化資料可在參考測站時空採樣均一的條件下產製。今年（105年）為計畫第三年，針對雨量資料的合理性查核分析，應用了陳等（2016）以雨量資料機率分佈為基礎來進行強降水事件判定值的分析，採用超越機率值（Exceedance Probability, EP）方法來協助檢測，藉由比對六個不同降水延時（Duration of rainfall）下，EP值為1%、5%的雨量閾值（threshold），作為判斷氣候極端降雨事件的標準，並配合比對天氣系統，判斷是否為合理值。

## 二、資料及研究方法

### (一) 資料來源

中央氣象局為有效掌握中小尺度天氣系統所引發之豪雨與監測豪雨，從1987年開始以逐步推動與執行建置計畫，布建台灣各主要流域的自動氣象站與自動雨量站，直至1998年始有東部區域的測站增設。為了全臺測站分佈密度的均勻性，因此設定測站選取條件有二：1.1998年至2012年共15年，2.資料齊整率需 $\geq 85\%$ ，選取符合條件之測站有110個氣溫目標站及318個雨量目標站。

氣溫資料採用於即時接收階段即已進行簡易檢核處理，依物理特性給定合理性範圍檢查大型錯誤檢核（Gross error check；又稱 Limit check）的原始資料，資料長度為1998年至2016年等近19年的逐時資料。

雨量資料則採用經過處理雨量特殊碼的QC（Quality Control）資料。根據二組雨量資料特殊碼定義，資料累積於後（-9996）為資料傳輸過程中，因訊號不良中斷時，會將中斷期間的所有雨量觀測值累計在訊號恢復後的第1筆觀測值中，已非當時那單一小時的降雨，而中斷期間的真實降雨狀況並無法準確得知。因此，參考陳等（2015）的處理方法，將累積於後的雨量值為正的資料視為缺值；並將累積於後的雨量值為0的連續特殊碼筆數，小於24筆的皆修正為0。資料長度同為1998年至2016年等近19年的逐時資料。

### (二) 研究方法

氣溫資料合理性查核分析，於本研究首先採用氣候背景範圍進行氣候檢驗（Climatological test），即是以氣候平均背景值來提取氣候離群值（Climate Outlier, CO）。另由於天氣的高變異特性，氣象記錄資料在特殊或劇烈天氣影響下，也常會表現出背離氣候離群現象。本研究鑑於單從氣候背景資訊所得離群值並不能明確區分確實錯誤與可能合理的資料，參考陳等（2015）藉由具可信度的參考推估值來檢測觀測數據的合理性，並運用迴歸方法由迴歸離群值（Regression Outlier, RO）來尋找可疑的錯誤數據。迴歸離群程度可用迴歸殘差表示，參考迴歸診斷分析常用指標，使用經標準化的刪除式殘差（或稱學生化殘差, Studentized Residuals），此殘差指標若符合  $t$  分佈，可幫助標準化定義合理值範圍（Yang and Xu, 2009）。再經由天氣系統比對，判斷是否為合理值。

雨量資料合理性查核分析原則，是參考陳等（2016）先將資料分為夏、冬兩季，其中夏季月份為4至9月，冬季月份為10月至次年的3月，計算318個雨量測站的觀測值超越機率值（Exceedance

Probability，以下簡稱EP值）。對於超越機率雨量閾值的計算乃參考陳與盧（2007）的發表。將經處理後的雨量資料以箱型計數（box-counting）方式重新組成六種延時的降雨事件（1hr、3hr、6hr、12 hr、24 hr和48 hr），根據線性動差法（L-moments）的推估結果採用皮爾森第三型分布（Pearson type III），模擬降雨事件雨量的母體機率分布，並計算EP值為1%和5%的雨量閾值作為判斷氣候極端降雨事件的標準。為避免小雨事件影響對超越機率閾值的估算，時雨量小於5mm/hr的資料不計。再經由比對天氣系統，判斷是否為合理值。

## 三、分析結果

本研究首先針對110個氣溫目標站，2011年1月至2016年6月的觀測資料，採用氣候離群值（CO）方法進行合理性查核分析。先以大於5SD（Standard Deviation）的氣候離群值為門檻值，列出所有可疑觀測資料筆數（表1），經由比對天氣系統，判斷是否為合理值。於近6年的551筆可疑資料中，大部份多為合理值，僅有3筆為錯誤值。當只看大於9SD時，有8筆可疑值，其中6筆皆有配合天氣系統，判斷為合理值；但2014年9月25日山佳站的24時（圖1）及2016年6月28日淡水站的16時（圖2）各有1筆資料，觀測值皆為0，資料明顯不連續，判斷為錯誤值。另1筆錯誤資料發生於2016年6月20日檳榔站的03時（圖3），為資料傳輸過程中受電波干擾，導致出現錯誤溫度值，該段受干擾期間資料應皆設為故障值。在年與年間的比較中，2013及2015年的資料中大於5SD以上的筆數較多，但大多為特殊天氣系統影響，經判斷皆屬合理值。根據最近6年氣候離群值與天氣系統比對的分析結果，造成比較大的合理離群值，主要是因為特殊天氣系統所造成，例如：颱風或其外圍環流造成氣流過山沉降增溫（焚風）、西南風增強的過山沉降增溫（焚風）、太平壓高壓增強、鋒面前緣暖區增溫造成的高溫現象；冬季大陸冷氣團系統、輻射冷卻、颱風降雨、西南氣流降雨、鋒面降雨、午後熱對流降雨造成的低溫現象。

再針對110個氣溫目標站，2011年1月至2016年6月的觀測資料，採用以UK（Universal Kriging）方法產製之參考值，利用迴歸離群值（RO）方法進行合理性查核分析。參考1998至2014年的氣溫RO分析結果（陳等，2015），觀測資料RO值大於10，幾乎皆為錯誤記錄。因此，初步先以RO大於10的離群值為門檻值，列出17筆可疑觀測資料（表2），經由比對每月天氣系統，判斷是否為合理值。於此17筆資料中，判斷有5筆為合理資料，主要為特殊天氣系統造成UK參考值的高估或低估，導致RO值有偏大的狀

況。另有12筆為錯誤資料，2016年6月28日淡水站(圖2)有1筆資料，觀測值為0，資料明顯不連續，判斷為錯誤值；其餘11筆資料皆為觀測儀器的穩定性不足所造成，其中9筆於採用氣候離群值合理性查核方法時，並無法被偵測出來。同時藉由檢查可疑資料的合理性，發現前後時間的觀測值亦有錯誤的情形，例如2015年6月14日的17時至17日的12時有連續資料錯誤的情形(圖4)。根據最近6年迴歸離群值與天氣系統比對的分析，統計造成上述合理離群值的特殊天氣現象，例如：日夜溫差大造成UK參考值的高估或低估；颱風環流過山沉降增溫(焚風)造成UK參考值的低估；局部降雨造成UK參考值的高估。

另針對318個雨量目標站，2016年1月至2016年6月的觀測資料，則採用以超越機率值方法來協助檢測，藉由比對六個不同延時下，EP值為1%、5%的雨量閾值，作為判斷氣候極端降雨事件的標準，並配合比對天氣系統，判斷是否為合理值。結果發現於2016年3月26日至27日鳳美站有出現異常大值的逐時雨量，最大值為217.5mm(圖5)，但當日並無明顯的降雨天氣系統，因此判斷該期間資料為錯誤值；另經由查詢二組該站的維護情形後，確認該站於3月22日至3月27日期間，自動雨量儀的磁簧開關故障，造成有異常的雨量值出現，此紀錄支持其雨量觀測值有高度異常的可能性，因此該站於此期間的資料應皆改為故障值-9999。經由此分析方法亦可找到破紀錄的極端降雨事件，例如2016年5月15日屈尺站於15時的雨量值為89mm(圖6)及2016年6月28日八德站於16時的雨量值為112.5mm(圖7)，兩者皆創下該站設站以來的整點雨量新紀錄。

#### 四、結語與討論

本研究透過分析1998年至2016年等近19年的110個氣溫目標站及318個雨量目標站之逐時資料齊整率，發現於近期的資料齊整率與早期建立時相比，有明顯的改善情形，主要為二組的自動測報系統汰換及新增計畫，大幅提高了自動站資料齊整率；另一方面，對於先前低齊整率的自動測站，二組亦透過進行資料補整作業程序，將正確可用資料填補回去，改善了整體資料齊整率。

自動測站氣溫資料合理性查核分析，分別導入了氣候離群值(Climate Outlier, CO >5)及迴歸離群值(Regression Outlier, RO >10)等2種方法來協助檢測，配合比對天氣系統，判斷是否為合理值。結果顯示該2方法確實可合理偵測極端高低溫事件，亦有找出多筆錯誤氣溫值的實例；雨量資料合

理性查核分析則採用超越機率值(EP)方法來協助檢測，藉由比對六個不同延時下，EP值為1%、5%的雨量資料閾值，作為判斷氣候極端降雨事件的標準，並配合比對天氣系統，判斷是否為合理值。結果顯示該方法確實可合理偵測出極端降雨事件，亦有找出多筆錯誤雨量值的實例。

透過以上3種合理性查核方法與天氣系統比對的分析，增加對於可疑資料合理性判斷的經驗累積，對於特殊天氣系統造成的極端事件可更加準確的判斷合理性，以利後續資料品管流程的改進；並可將偵測出的錯誤資料即時修正原始資料，以增進觀測資料的正確性使用。

#### 參考文獻

- 陳佳正、盧孟明，2007年：“臺灣極端降雨氣候事件判定方法”，大氣科學。35期第2號，105-117。
- 陳品好、沈里音、陳雲蘭，2015：“「臺灣長期氣候資料整合分析」計畫研究(2) —自動站雨量累計於後記錄值的問題分析及處理”，天氣分析與預報研討會，中央氣象局。
- 陳品好、陳雲蘭、沈里音，2016：“中央氣象局所屬雨量站降水超越機率值分析”，天氣分析與預報研討會，中央氣象局。
- 陳雲蘭、薛宏宇、呂致穎、陳品好、詹智雄、沈里音，2015：“「臺灣長期氣候資料整合分析」計畫研究(1) —自動氣象站長期氣溫觀測值合理性檢測方法探討及分析”，天氣分析與預報研討會，中央氣象局。
- Yan, X. and Su, X.: Linear Regression Analysis: Theory and Computing, World Scientific, Singapore, 2009.

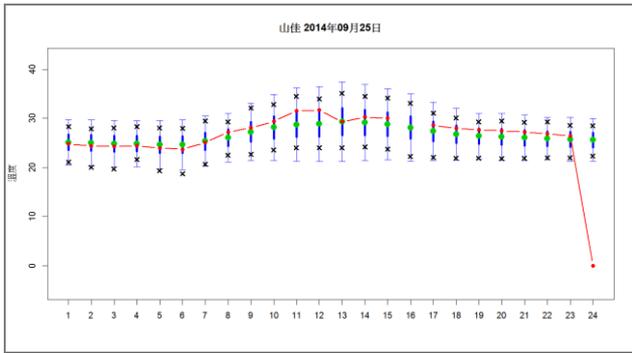


圖1 2014年9月25日山佳站氣溫逐時序列圖

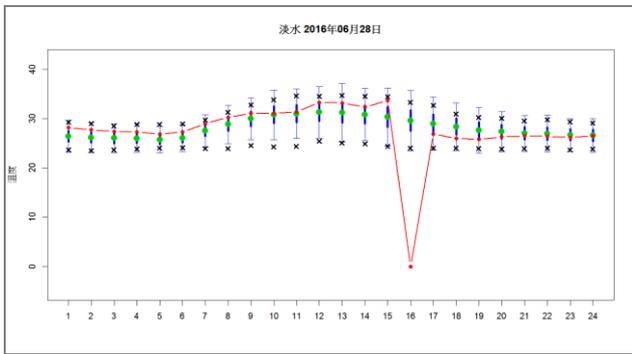
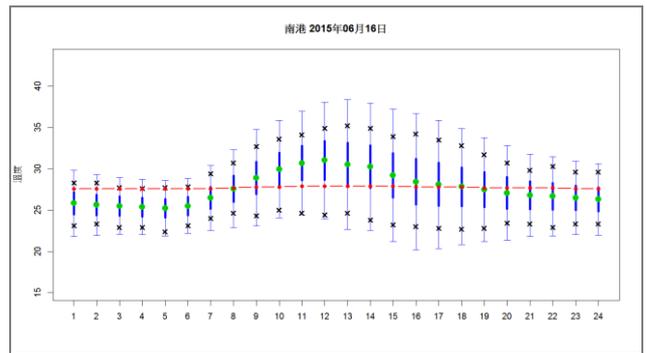
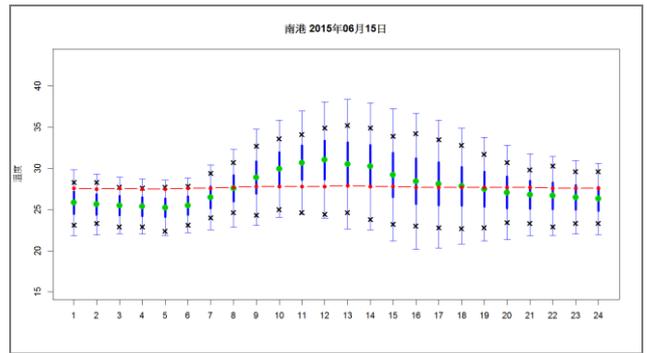
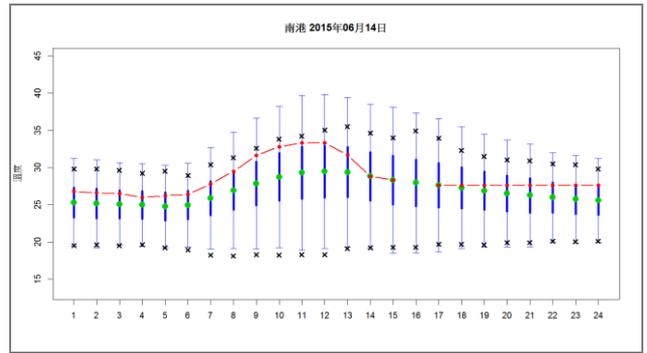


圖2 2016年6月28日淡水站氣溫逐時序列圖

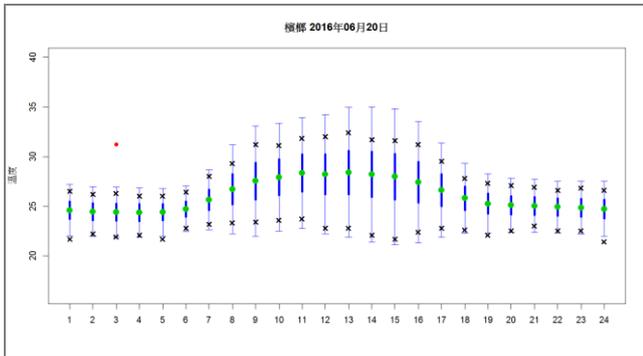


圖3 2016年6月20日檳榔站氣溫逐時序列圖

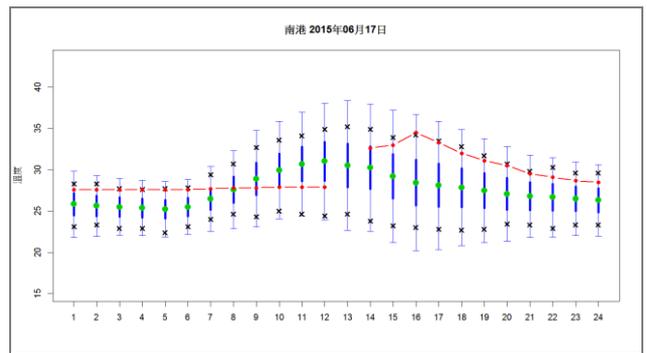


圖4 2015年6月14日至17日南港站氣溫逐時序列圖

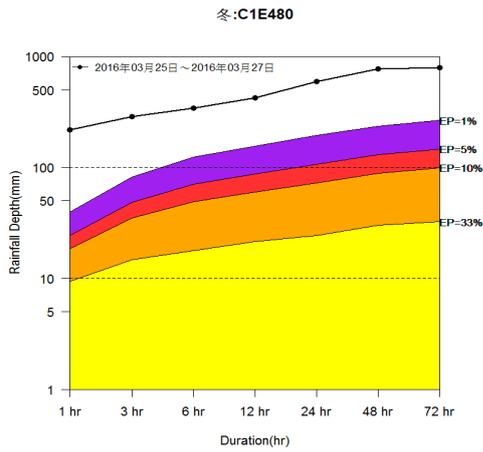


圖5 2016年3月25日至27日鳳美站各延時EP分析圖

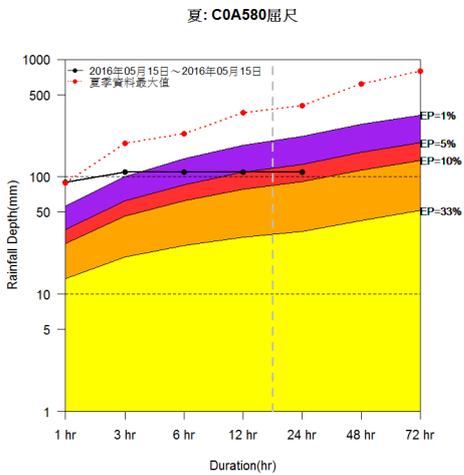


圖6 2016年5月15日屈尺站各延時EP分析圖

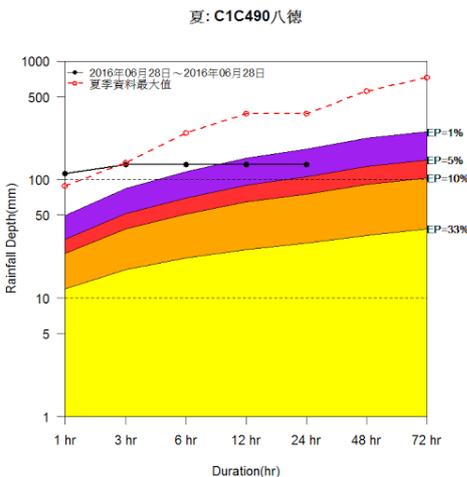


圖7 2016年6月28日八德站各延時EP分析圖

表1 2011年至2016年氣候離群值筆數統計 (紅字為錯誤值)

	> 5SD	> 6SD	> 7SD	> 8SD	> 9SD	> 10SD
2011	17	0	0	0	0	0
2012	24	5	0	0	0	0
2013	165	45	21	9	5	3
2014	45	11	5	1	1	1
2015	233	95	33	7	1	0
2016	67	10	1	2	1	1

PS. 2016年資料僅統計至6/30止

表2 2011年1月至2016年6月迴歸離群值筆數統計 (紅字為錯誤值)

	1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月
2011	0	0	0	0	0	0	0	0	0	0	0	0
2012	0	0	0	0	0	0	0	0	0	0	0	0
2013	0	0	0	2	0	0	0	0	0	0	0	0
2014	0	0	7	0	0	0	0	0	1	1	0	0
2015	0	0	0	0	0	2	0	1	0	0	0	0
2016	0	0	0	0	1	2	—	—	—	—	—	—