

即時雨量資料品質檢覈

鄭安孺² 李天浩¹ 顧欣怡³ 高慧萱² 陳怡彤²

國立臺灣大學水工試驗所¹

多采科技有限公司²

交通部中央氣象局³

摘要

時雨量空間分布與時變性均大，不易建立可靠的檢覈機制。本研究根據中央氣象局台灣西部自動雨量站系統採用「事件觸發(event triggering)」與「定時確認」組合通訊邏輯的特性，提出應用普通克利金法(Ordinary Kriging technique)篩檢因為傳輸失敗而「累積於後」的超大累積雨量值以及因為雨量站故障而始終回傳「零雨量」的錯誤雨量檢覈機制。

「累積於後」的異常雨量值，除了其值將普遍高於同時間其他正常測站雨量記錄外，另外伴隨的訊息是該站前一時間的記錄為負值的錯誤代號。雨量站故障的異常「零雨量」值，除了同時間其他正常測站均有雨量記錄而該站雨量記錄為零外，另外伴隨的訊息是前一時間的雨量記錄亦為零。對於豪大雨事件，若測站觀測雨量與克利金法的估計雨量的誤差值大於六倍平均估計誤差，且觀測條件滿足前述錯誤訊息時，雨量可能為異常。

以 2005 年至 2010 雨量資料進行分析，研究結果顯示，本研究所提出的雨量資料檢覈法可有效篩檢出因為傳輸失敗而「累積於後」以及因為雨量站故障而始終回傳「零雨量」的兩類錯誤雨量。

一、前言

雨量觀測對於民生用水、農業灌溉、工業發展、乃至於防災策略的擬定是必要的資訊，即時降雨觀測對於防救災尤其重要。目前包括中央氣象局在內等各政府單位於臺灣地區已建置超過八百座以上的雨量站，可於降雨事件發生時即時量測降雨量，並且回傳至中央氣象局。但即時回傳資訊有可能會因為連線中斷或異常的問題而產生錯誤，以至令使用者對於情勢產生誤判。因此有必要發展雨量資料檢覈機制，可於接收到雨量訊息時，即時篩選出錯誤資料。

雨量資料的檢覈是一件困難的事，主要原因是臺灣地區雨量空間分布與時變性均大，因此不容易建立可靠檢覈機制。本研究根據中央氣象局資料傳輸的邏輯，認為對於即時雨量資料的檢覈，應著重在篩檢出因為傳輸失敗而「累積於後」的超大累積雨量值以及雨量站故障而始終回傳「零雨量」的錯誤資料。在此前提下，本研究乃發展出應用克利金法(Kriging technique)篩檢「累積於後」及「零雨量」兩種錯誤雨量資料之即時雨量檢覈法。

二、雨量資料品質檢覈方法

中央氣象局台灣西部自動雨量站系統在建站時的

系統設計，是採用「事件觸發(event triggering)」與「定時確認」的組合通訊邏輯。

「事件觸發」指的是當雨量站的傾斗式雨量筒因為降雨至少發生一次滿斗傾倒時，由測站端主動啟動通訊，透過中繼站等將觀測值傳送到中央氣象局。如果測站和中央氣象局之間的通訊沒有成功，在中央氣象局端，因為 10 分鐘或 1 小時內沒有收到某測站傳來訊號，便認為該測站該 10 分鐘或 1 小時的降雨量為 0，對外傳送該測站的即時降雨記錄為 0，但實際降雨量觀測值未必等於 0；在測站端的處理方式，是將未成功傳遞的降雨量「累積於後」，待下次成功連結到中央氣象局時，將該時段以及之前連續通訊不成功、「累積於後」的所有降雨量，當成單一數值送出。當大面積同時發生高強度降雨時，中繼站區域內多個雨量站會因為同時密集嘗試傳送資料，而致部份測站無法有效連接時，便可能在降雨當時無法傳輸。

「定時確認」指的是每個雨量測站不論是否有降雨，都必須在設定的時間間隔內(約 3 小時)，透過中繼站與中央氣象局聯繫，以確認該站為正常運作中；如果超過設定的時間間隔測站仍未聯繫，即表示該站可能已經發生故障，中央氣象局的處理方式，是將該站的即時觀測記錄改以「-98」代號表示。

「累積於後」的異常雨量值，除了其值將普遍高於同時間其他正常測站雨量記錄外，另外伴隨的訊息

是該站前一時間的記錄為「-98」(即時觀測)或「-9996」或「-9999」(歷史記錄)的錯誤代號。雨量站故障的異常「零雨量」值,除了同時間其他正常測站均有雨量記錄外,另外伴隨的訊息是該站前一時間的記錄亦為「零雨量」。根據前述之特性,本研究設計雨量檢覈的程序為,可以先做最基本的合理範圍檢覈,然後使用普通克利金法做空間分布合理性檢覈。

茲說明研究方法如下。

(一) 雨量合理範圍檢覈

雨量資料檢覈的第一步為檢查資料是否在合理範圍內。圖 1 為統計西元 2005 年至 2010 年 558 個測站的歷史資料所得各測站的最大時雨量(紅色+字)與次大時雨量記錄(藍色)分佈,橫軸為測站數,縱軸為降雨量。以人工方式檢視其中最大時雨量超過 200mm/hr 以上的記錄前後時間的資料,可以發現這些記錄應均為「累積於後」的錯誤。篩除這些資料後,可發現 2005 年至 2010 年間最大降雨事發生在 2009 年 08 月 10 日 02 時莫拉克颱風期間,中央氣象局 C0A950 鼻頭角站,其值為 189.0mm/hr。統計結果顯示,時雨量超過 200 mm/hr 的可能性極低,因此將時雨量的上限門檻取為 200mm/hr+10%,即 220mm/hr。

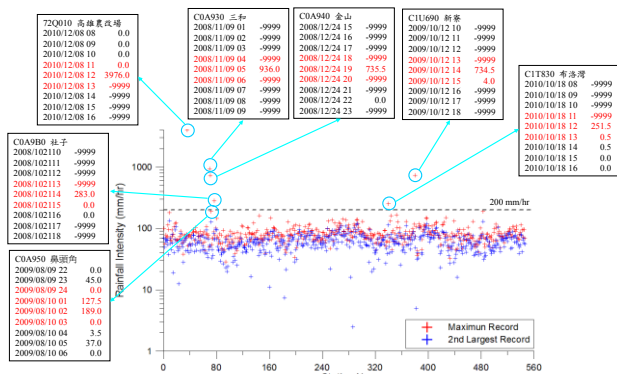


圖 1 根據 2005 至 2010 年資料統計得各雨量站最大與次大降雨記錄

(二) 克利金法空間分布合理性檢覈

克利金法(Kriging technique)又名地質統計法(Geostatistics technique)是根據資料在空間中分佈的統計特性,決定線性內插係數的一種技術(Bras and Rodriguez-Iturbe, 1985;Kitanidis, 1993)。氣象局於 2009 年「發展成鄉鎮逐時天氣預報」計畫中(中央氣象局, 2009),曾對應用普通克利金法建立高解析度網格點雨量場進行遮蔽測站試驗(Observation System Simulation Experiment)效果評估,研究結果顯示普通克利金法可適用於內插雨量資料。

利用克利金法進行雨量資料檢覈的作法,由鄰近雨量站觀測值估計每個測站的降雨量和信賴區間,若觀測資料值落在信賴區間之內,則通過檢覈;若觀測資料值落在信賴區間之外,且滿足且符合「累計於後」或「零雨量」異常站的特性者,則給予異常檢覈代碼,並且在後續進行網格雨量內插時,不使用該雨量站的資料。

克利金法的估計步驟是分爲二部份:

1. 結構分析—由歷史或觀測資料,迴歸統計協變異數(covariance)隨距離變化的情形,決定「變異圖」或「半變異圖」(variogram or semi-variogram)函數。
2. 最佳線性不偏估估計(Best Linear Unbiased Estimation-BLUE)—假設估計值為已知值的線性權重平均,根據不偏估和最小估計誤差變異數兩項原則,根據變異圖或半變異圖導出協變異數矩陣,進而推估各測站對估計點的權重係數值。

實際檢覈方法說明如下。

1. 雨量資料結構分析

進行資料檢覈前,必須要事先準備克利金法所需半變異圖。「半變異圖」是觀測樣本資料的協變異數隨距離變化情形。令 $\mathbf{u} = u(x, y)$ 為空間位置座標, $Z(\mathbf{u})$ 為空間隨機變數,則變異圖定義為:

$$2\gamma(d_{ij}) = E\left[\left(Z(\mathbf{u}_i) - Z(\mathbf{u}_j)\right)^2\right] \quad (1)$$

即距離等於 d_{ij} 的兩兩組合變數差異平方的「期望值」。 $\gamma(d_{ij})$ 稱為半變異圖。若隨機場為定常性,亦即資料之統計特性不因測站位置而有差異,則半變異圖 $\gamma(d_{ij})$ 和協變異數 $\text{cov}(\mathbf{u}_i - \mathbf{u}_j) = \text{cov}(d_{ij})$ 為互補函數

$$\gamma(d_{ij}) + \text{cov}(d_{ij}) = \sigma^2 \quad (2)$$

其中, σ^2 為定常性空間變數 $Z(\mathbf{u})$ 的變異數。

以克利金法進行空間雨量場內插時,是以同時間所有雨量站的觀測雨量即時迴歸統計半變異圖函數;但是以克利金法進行雨量檢覈時,因為只是做定性分析,因此不需要逐時進行迴歸,可使用歷史資料迴歸事先統計通用的半變異圖函數。除可以加速即時資料檢覈時效外,亦有助於資料分析的穩定性。但是因為考慮到各季節降雨的成因不同,因此將資料分爲五季,分別是:12-2 月、3-4 月、5-6 月、7-9 月以及 10-11 月,取各季所含月份的加上前後 10 天的資料,分季迴歸統計半變異圖函數。各季資料加上前後 10 天的意義,是減少人為分季對分季的交接時節資料的連續性檢覈造成影響。

由於各時間點的降雨量強度不同,因此需要將雨量資料正規化,再行分析。分析步驟如下:

- (1) 逐時計算各時間所有測站雨量資料(不含缺漏以及零雨量資料)的平均值:

$$\mu_t = \sum_{i=1}^{n_t} p_{i,t} / n_t, \quad p_{i,t} > 0, \forall i$$

其中, μ_t 為 t 時間的平均雨量; n_t 為 t 時間的雨量資料數。 $p_{i,t}$ 為 t 時間、 i 站的雨量觀測值。

- (2) 逐時求取各時間所有雨量資料(不含缺漏以及零雨量資料)的變異數:

$$\sigma_t^2 = \sum_{i=1}^{n_t} (p_{i,t} - \mu_t)^2 / (n_t - 1)$$

其中, σ_t^2 為 t 時間的雨量變異數。

- (3) 將雨量資料除以該時間的變異數作正規化,即令

$$z_{i,t} = p_{i,t} / \sigma_t$$

- (4) 用 $z_{i,t}$ 資料求取半變異圖,半變異圖使用含金塊效

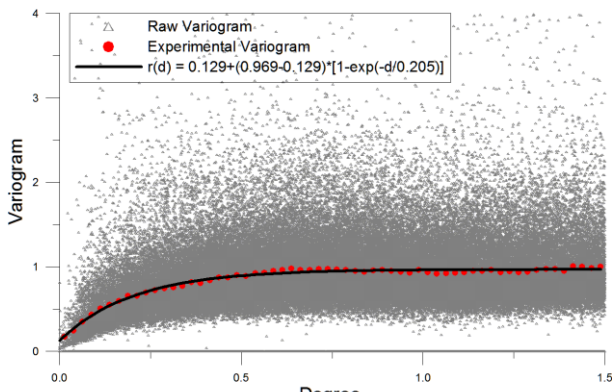
應 C_0 的指數模型：

$$\gamma(d_{ij}) = C_0 + \left[\sigma_*^2 - C_0 \right] \left[1 - \exp\left(-\frac{d_{ij}}{L}\right) \right] \quad (4)$$

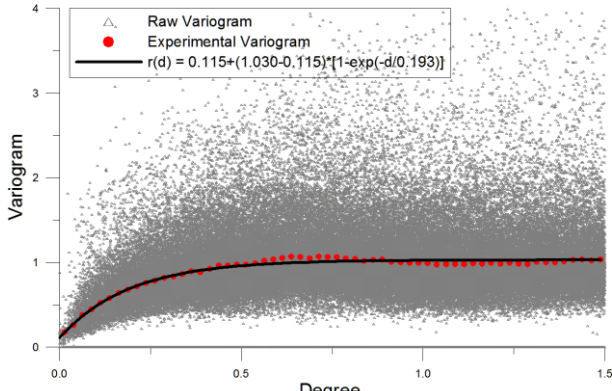
其中， σ_*^2 是正規化雨量的變異數，理論上應該接近 1； C_0 為金塊效應(nugget effect)，所反映的訊息是，即使是很近距離的兩點，也可能存在的差異； L 是影響長度(integral length)，反映距離對相關性的影響，當兩測站距離大過 L 時，兩站間的相關性已經降到 0.37 以下；當兩測站距離大過 $1.5L$ 時，兩站間的相關性將降到 0.22 以下。

圖 2 所示為以 2005 年至 2009 年降雨資料，所統計得各季的正規化半變異圖。其中灰色 Δ 符號為觀測樣本資料的原始半變異數；紅色 \bullet 實心圓符號所示為實驗半變異數；實線所示為以實驗半變異數所套配的半變異圖函數。如圖所示，等距離的原始半變異數在不同的時空點可能會有很大的差異。因此在進行迴歸之前，通常會將距離軸分為幾個區間，然後將同區間內距離相近半變異數取平均求得實驗半變異數，由實驗半變異數就可以看到在排除了局部干擾的因素下，兩測站的相關性隨距離變化的情形，並且可以用一簡單指數函數來描述這種關係。

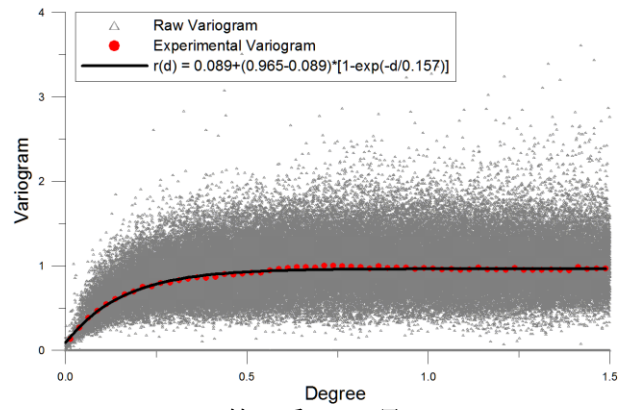
不同季節的正規化半變異圖差異並不大，變異數接近於 1，影響半徑則 L 介於 0.15 ~ 0.22 度，約為 16~24 公里之間，其中以 5-6 月的影響半徑最小，反映春、夏間月是小尺度的降雨。



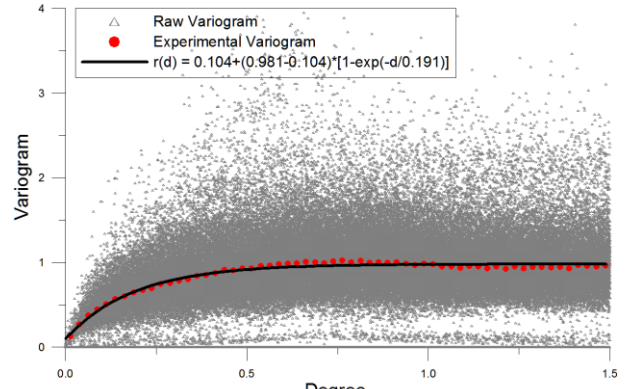
第一季：12-2 月



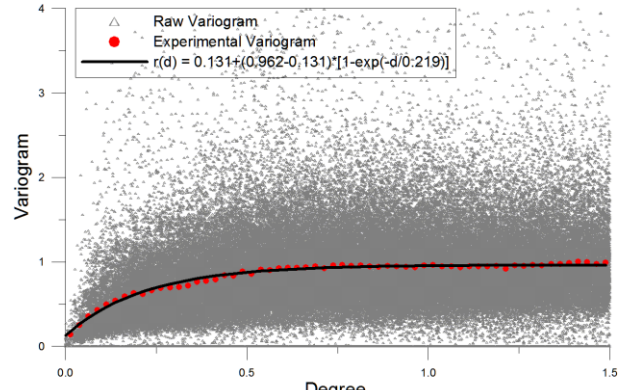
第二季：3-4 月



第三季：5-6 月



第四季：7-9 月



第五季：10-11 月

圖 2 根據 2005 至 2010 年時雨量資料統計得之正規化半變異圖

2. 資料檢覈

由於雨量檢覈的目標是要篩檢出「累積於後」以及測站故障「零雨量」異常資料，因此將只對強降雨事件進行檢覈。事件的篩選標準是，同時間的雨量記錄中，至少有 1 筆資料是超過 20mm/hr。檢覈步驟如下：

(1) 計算該時間雨量不為零的「有效」時雨量平均值和變異數：

$$\mu_t = \frac{\sum_{i=1}^{k_t} p_{i,t}}{k_t} ;$$

$$\sigma_t^2 = \frac{\sum_{i=1}^{k_t} (p_{i,t} - \mu_t)^2}{(k_t - 1)} , p_{i,t} > 0, \forall i$$

k_t 為不為零的有效時雨量站數。

- (2) 選取各雨量站的內插參考站。首先以雨量站 i 為中心，令選取半徑為 $d_0 = 0.3$ 度，選取距離 $d_{ij} < d_0 = 0.3$ 範圍內的有效雨量(含零雨量)，如果搜尋到的站數 $m_{t,i}$ 小於 30 站，則將搜尋半徑擴大 0.05 度，如此逐次擴大半徑，直到搜尋到至少 30 座參考站為止；
- (3) 採用普通克利金法，內插估計第 i 站的雨量值 $\hat{p}_{i,t}$ 。令，

$$\hat{p}_{i,t} = \sum_{j=1}^{m_{t,i}} \lambda_j p_{j,i} \quad (4)$$

其中， λ_j 為第 j 個鄰近測站的權重係數； $p_{j,i}$ 為第 j 個鄰近測站的雨量觀測值。根據不偏估條件為：估計值的期望值應等於平均值，即

$$E[\hat{p}_{i,t}] = E\left[\sum_{j=1}^{m_{t,i}} \lambda_j p_{j,i}\right] \Rightarrow m = m \sum_{j=1}^{m_{t,i}} \lambda_j$$

因此

$$\sum_{j=1}^{m_{t,i}} \lambda_j = 1 \quad (5)$$

估計誤差變異數為 $L = E[(\hat{p}_i - p_i)^2]$ ，根據最小估計誤差原理，使估計誤差變異數為最小的必要條件是： L 對任意權重係數 λ_j 的微分值必須為零，

$$\frac{\partial L}{\partial \lambda_j} = 2 \sum_{k=1}^{m_{t,i}} \lambda_k \text{cov}_{jk} - 2 \text{cov}_{ik} = 0 \quad (6)$$

利用 Lagrange multiplier， ν ，結合以上(6)式「估計誤差變異數最小化條件」以及(5)式「不偏估條件」：

$$\frac{\partial L}{\partial \lambda_j} + 2\nu \left(\sum_{i=1}^n \lambda_i - 1 \right) = 0$$

可推導出普通克利金法協變異數矩陣方程式為：

$$\begin{bmatrix} \sigma_i^2 & \text{cov}_{1,2} & \cdots & \text{cov}_{1,m_{t,i}} & 1 \\ \text{cov}_{2,1} & \sigma_i^2 & \cdots & \text{cov}_{2,m_{t,i}} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \text{cov}_{m_{t,i},1} & \text{cov}_{m_{t,i},2} & \cdots & \sigma_i^2 & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_{m_{t,i}-1} \\ \nu \end{bmatrix} = \begin{bmatrix} \text{cov}_{i,1} \\ \text{cov}_{i,2} \\ \vdots \\ \text{cov}_{i,m_{t,i}} \\ 1 \end{bmatrix} \quad (7)$$

上式中的 σ_i^2 為該小時所有大於 0 的「有效」資料的變異數，同時，

$$\text{cov}_{ij} = \sigma_i^2 \left[\sigma^{*2} - \gamma(d_{ij}) \right] \quad (8)$$

- (4) 計算估計點的估計誤差變異數：

$$s^2 = \sigma_{\hat{p}_{i,t}}^2 = \left(\sigma_t \times \sigma^* \right)^2 - \nu - \sum_{j=1}^m \lambda_j \text{cov}_{i,j} \quad \text{for } \forall i \quad (9)$$

- (5) 設定信賴區間為 6 倍的估計誤差值。當 $|\hat{p}_{i,t} - p_{i,t}| < 6s$ ，接受觀測值為真，否則觀測值可能有誤，進一步檢查：

- 若 $\hat{p}_{i,t} < p_{i,t}$ ，且 $p_{i,t-1} < 0$ ，屬於「累積於後」的異常狀況。
- 若 $\hat{p}_{i,t} > p_{i,t}$ ， $1 > p_{i,t} > 0$ 且 $1 > p_{i,t-1} > 0$ ，屬於「零雨量」的異常狀況。

不滿足以上兩項條件之一的狀況，即使估計與觀測的絕對差值大於 $6s$ ，仍接受觀測值為真。由於考慮到如果參考站的觀測資料有誤時，可能會因為錯誤資料的汙染，反而將測站正常的觀測資料濾除，因此實際在進行資料檢覈式，是對所有的站先逐站計算估計誤差比值，令

$$d_i = \frac{|\hat{p}_{i,t} - p_{i,t}|}{6s} \quad (10)$$

若 $d_i > 1$ ，則先將該站標示可能異常站，當所有的站都測試完畢，比較各異常站中，將其中 d 值最大的異常站剔除，然後再對其他的異常站重新檢覈，再次檢覈時不會使用被剔除的異常站為參考站，如此重複檢覈直到異常站數量為 0 為止。

以上步驟(5)的信賴區間若取的太窄，可能會犧牲掉部分正常的觀測，若取得太寬，則可能無法篩檢掉錯誤的資訊。假設觀測值與估計值的誤差接近於常態分佈，當顯著水準取為 $\alpha = 0.1\%$ ，且樣本數趨近無限大時， t 分布值約為 $t_{\alpha/2} \approx 3.1$ 。但考慮雨量空間分布與時變性均大，誤差為常態分布的假設亦不一定成立，因此取更寬鬆的信賴區間。6 倍的估計誤差值是本研究經測試認為較合理的係數。

表 1 所示為檢覈代碼說明。通過檢覈的雨量資料給予代碼 1；「零雨量」的異常狀況給予代碼 31，「累積於後」的異常狀況則是給予代碼 61。無法進行檢覈，例如篩選到參考站數不足 2 站，或未進行檢覈的狀況，例如同時間各測站的雨量均小於 20mm/hr，則給予代碼 0。

本研究提出的資料檢覈法僅適用於強降雨事件：同時間的雨量記錄中，至少有 1 筆資料是超過 20mm/hr。因為如果所有站的雨量都不大時，則可能因為站與站的雨量差異不大，導致誤判的情形發生，尤其是對於「零雨量」型錯誤的研判，可能會將真正無雨量的資料誤判為是錯誤資料。

表 1 為檢覈代碼說明。

檢覈代碼	意義	檢覈結果
-99	t時間無觀測值	無法檢覈
0	無法檢覈或未檢覈	未檢覈
1	通過合理性檢覈	通過
31	「零雨量」的異常狀況	未通過
62	「累積於後」的異常狀況	未通過

三、分析結果

取 2005 年至 2010 年逐時雨量進行檢覈，檢覈結果如圖 3 所示，圖中顯示，其中，2005 年水利署 01A190

碧湖站有多筆「零雨量」的異常值。經檢查，發現有多筆「零雨量」異常資料是發生集中在 2005 年 08 月 31 日。將 2005 年 08 月 30 日 01 時至 9 月 1 日 24 時 01A130 碧湖站與鄰近氣象局 C0A540 四堵站以及 C0A530 坪林站雨量資料取出，繪製時間序列圖，其結果圖 4 所示。圖中所圈選的時段為碧湖站被篩檢出有零雨量異常的資料，而同時段四堵站及坪林站觀測到有降雨發生，因此碧湖站的「零雨量」錯誤檢覈結果應為可信。

圖 5 為一「累積於後」錯誤資料的範例。圖中所示為 2005 年 05 月 12 日 01 時至 05 月 14 日 24 時氣象局 C0M410 馬頭山站、C1M540 小公田站、以及 C0X020 關子嶺站降雨時間序列比較，其中，馬頭山站於 05 月 13 日 07 時至 13 時間共六小時無資料，而同時段除 09 時外，其他時間小公田站與關子嶺站均有雨量記錄，由此可推測，此時段馬頭山站降雨機率應該很高。14 時小公田站與關子嶺站時雨量分別是 21 以及 23mm/hr，但馬頭山站雨量記錄卻是高達 91mm/hr，因此馬頭山站 14 時被判定為「累積於後」錯誤檢覈結果應為可信。

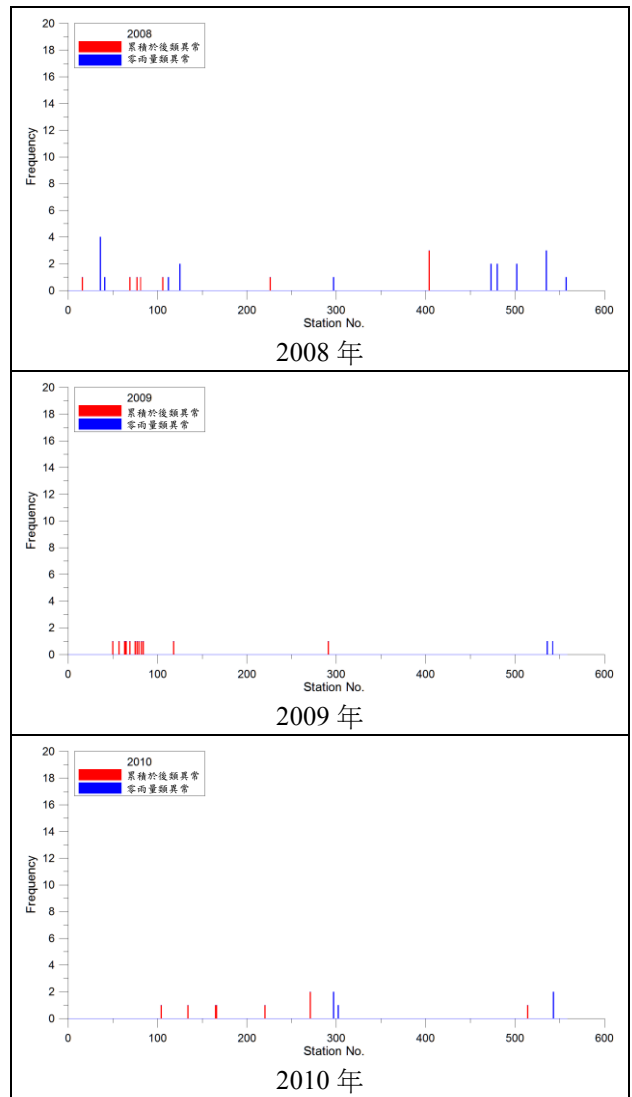
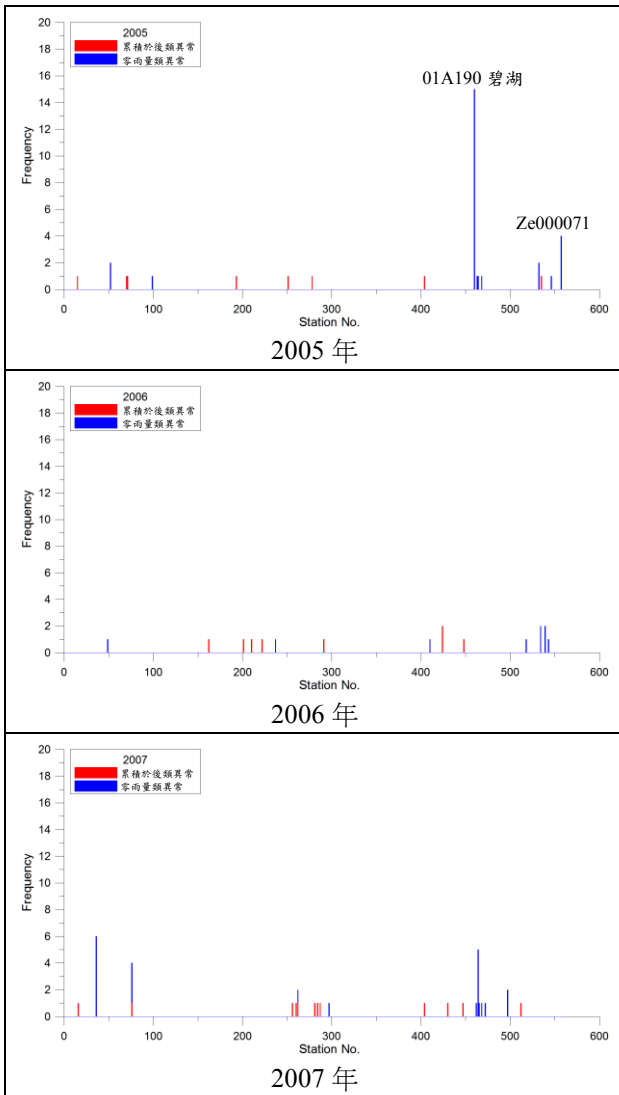


圖 3 2005-2010 年時雨量資料檢覈結果

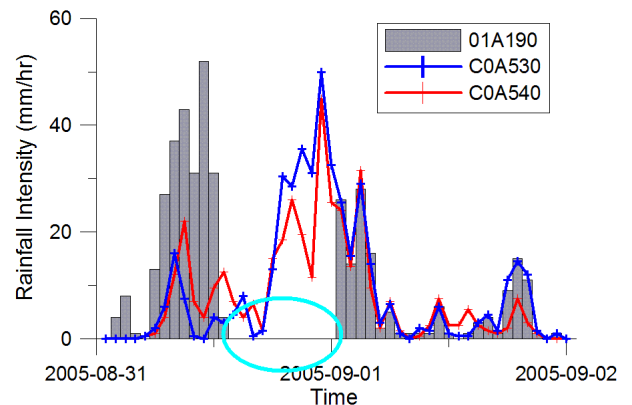


圖 4 2005 年 08 月 30 日 01 時至 09 月 01 日 24 時 01A190 碧湖站與 C0A540 四堵站及 C0A530 坪林站降雨時間序列比較

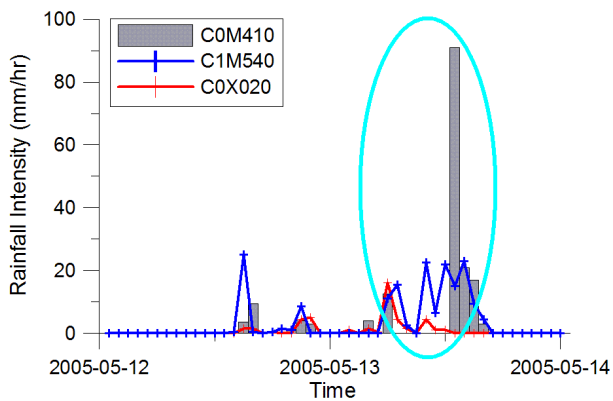


圖 5 2005 年 05 月 12 日 01 時至 05 月 14 日 24 時
COM410 馬頭山站與 C1M540 小公田站及 C0X020 關
子嶺站降雨時間序列比較

四、結論

雨量資料空間分布與時變性均大，因此品質檢覈不易。本研究根據中央氣象局即時資料，「事件觸發 (event triggering)」與「定時確認」的組合通訊邏輯。歸納因為傳輸失敗而「累積於後」以及因為雨量站故障而始終回傳「零雨量」兩類雨量錯誤類型的伴隨訊息，提出以克利金法為基礎之即時雨量檢覈法。2005 年至 2010 年雨量資料檢覈結果顯示，本研究所提出的雨量資料檢覈法可有效篩檢出強降雨事件的「累積於後」以及「零雨量」兩類錯誤資料。

五、參考文獻

- 交通部中央氣象局，”應用克利金法建立高解析度網格點氣象數據研究”期末報告書，2009。
- Bras, R. L. and Rodriguez-Iturbe, I.:” Random functions and hydrology, Addison Wesley”, Reading, Mass. , 1985.
- Kitanidis, P.K., “Gestistics” in “Handbook of Hydrology” edited by D.R. Maidment, McGraw Hill. , 1993.