

# 模組化之 Storm Tracker 資料校正流程建置

吳閔倫 張巧薇 蘇世顯\* 劉清煌 楊菁華 陳冠儒

中國文化大學

## 摘要

Storm Tracker 微型探空儀 (ST) 在性價比上有比現階段作業化之探空儀有更好的表現，具備可進行高密度、高頻率觀測的優勢，是對低層大氣或邊界層進行觀測研究時的極佳工具。由於 ST 是新開發之觀測設備，所以在使用觀測資料前需要進行詳細的誤差校正，以期獲得具有科學研究水準之資料品質。本研究將介紹針對 Storm Tracker 所發展之校正流程，透過此一標準化之資料後處理程序，可以使觀測資料之格式統一且與現階段作業化標準探空儀具有相同品質之探空資料。

根據尤等在 2020 年的研究指出，ST 在 2016~2019 年夏季的雙北暴雨實驗 (TASSE) 中被大量的使用進行邊界層觀測，在實驗中與氣象局標準作業使用的 Vaisala RS-41 探空儀進行了超過 500 次的共同施放實驗，並收集了超過 40 萬筆的溫度、濕度及壓力資料進行對比及校正。後續在 2020 年宜蘭劇烈降雨觀測實驗 (YESR2020) 中，嘗試利用與該研究相同的資料校正方式，但在校正過後仍存在溫度較低且濕度較高的偏差。為此，本研究將利用機器學習模型比對共同施放的資料組，並納入比濕、輻射量等物理量進行模型訓練 (training)，目標是修正 ST 在水氣量及溫度測量上的誤差，並且透過將流程模組化，使資料校正標準統一且快速，有利於觀測後分析與統計。

ST 資料組校正流程共分為五個階段，每一個部分皆模組化為獨立的模塊。第一階段是將 L0 原始資料格式齊制化與檢查資料時間排序，並且利用氣壓連續下降作為氣球升空之判斷依據，自動化的辨識升空時間。除自動化判定的升空時間外，也可以將觀測時的人工紀錄輸入，以篩除未升空前(地表校驗階段)之觀測資料，並產製包含升空標記的 L1 資料。第二階段將資料轉換為資料處理過程之固定格式，只保留部分校正模型所需資料，並自動化濾除觀測資料中之缺失值 (missing value)，在此模組中將計算校驗過程中所需的額外物理量，包含水平風場 (U, V)、比濕 (qv) 與地表晴空輻射量。這些資訊都將輸入第三階段的校正模型中進行校正，並產製出校正後的 L3 資料。第四階段則是將校正後資料內插至相同高度間距 (1 hPa)，並產製可供比對之 L4 資料。此外，在流程中我們獨立發展了銜接 NCAR-ASPEN 做後處理的模塊，可利用 ASPEN 進行資料過濾，並轉換為大部分觀測資料通用之 EOL 格式，這屬於整個流程的第五個部分。

整體來說，原始的觀測資料，經過五個步驟的資料處理流程，將輸出格式統一的原始資料及校正後資料共 6 個，其中包含各階段對資料進行的調整紀錄以及初步統計，可提供使用者快速瞭解資料特性。此流程能將探空資料以統一格式進行輸出輸入，並且每一個模塊包含作為核心的資料校正模型可以隨意替換，未來可發展具備可對應不同觀測環境條件和資料後處理需求的特化校驗模型，並可快速進行模型之間的校正能力比較。這對於未來發展新的 ST 校驗核心，並透過比對選出最適合的資料校正程序，保有極高的彈性。這樣的流程包含其他的優勢有：1. 具備高模組化的程式可攜性，允許在不同作業系統甚至小型虛擬平台上運

行及維護；2.高效能校驗速度，可在資料接收後立即做出校正並提供資訊；3.資料輸出為一般通用檔案格式，可滿足各種後處理的軟體需求，使觀測後的資料校正過程更加流暢且一致。這部分工作未來將由文化大學水文資料庫提供相關的服務，提供使用 ST 進行野外觀測實驗的人員進行資料即時校正並且建檔分析。未來資料校驗流程中，將陸續納入地面測站資料校正儀器誤差(ground check)、非線性 AI 校驗模組、即時觀測誤差查核統計模組與 ST 資料分析套件等功能。

**關鍵字：**探空觀測、機器學習、探空校正、校正流程

## 一、前言

過去幾年間，由台大林博雄老師團隊研發的新型探空儀器 – Storm Tracker (ST)由於較低的施放成本及輕便的特性，開始被應用在許多的觀測實驗中。ST 相較於氣象局作業用的 Vaisala RS-41 探空而言，在需要高頻率和高密度的邊界層觀測中，在實驗上具有有明顯的優勢。例如 2016-2020 由郭鴻基教授主持的雙北暴雨實驗(TASSE)和 2020 年 11 月由劉清煌老師主持的宜蘭劇烈降雨觀測實驗 (YESR2020)中，皆被大量的應用在邊界層觀測上，為觀測實驗提供了高時間與高空間解析度的近地面大氣溫度剖面資料。

先前的研究中，Ciesielski et al. 在 2012 年發表的論文提出了基本的探空資料校正流程概念，其中包括資料格式統一、經過初步移除誤差值、透過統計或其他方式針對環境特性進行校正誤差，最後以對使用者較為友善的資料格式輸出。針對誤差校正部分，Ciesielski et al 在 2009 年針對北美季風實驗 (North American Monsoon Experiment, NAME)使用的 Vaisala RS80-H 建置可供對比的溫濕度校正表，其概念是利用累積分布函數(Cumulative Distribution Function, CDF)做統計特性的比較，並藉此來修正誤差。

而尤等在 2020 年的研究中，利用 TASSE 期間，ST 與氣象局標準作業使用的 Vaisala RS-41 探空儀進行超過 500 次的共同施放資料，建置針對 ST 的溫濕度校正表。後續在 2020 年 YESR2020 期間，嘗試利用與前面提到的 CDF 校正表進行資料校正，但在校正後仍存在著溫度較低且相對濕度較高的偏差(圖一)。為此，我們將以本實驗室開發的機器學習模型為核心(Chang et al., 2021)，建置一個高度模組化的 ST 資料校正流程，並擴展其他應用模組提

供給觀測人員能夠在第一時間快速取得校正資料進行初步分析。

## 二、資料校正流程

整個資料校驗流程總共可分為 6 個步驟，並且將經過處理之資料分成 4 個不同的階段(圖二)。利用機器學習模型比對共同施放的資料組並納入比濕、輻射量等物理量進行模型訓練(training)，目標是修正 ST 在水氣量及溫度測量上的誤差。

### (一) 自動化判定升空

第一階段以 ST 的原始資料作為輸入，由於 ST 施放前會在地面待命，當接收穩定後再施放，故原始資料中會含有施放之前的數據。在第一階段中以觀測氣壓連續下降十筆資料作為探空氣球升空的依據，自動化辨識實際升空時間，並將資料標記為已升空資料與未升空資料，以利後續使用。若遇到施放失敗或是數據接收出現錯誤導致無法判定時將輸出至 L1 的紀錄中。

### (二) 計算機器模型校正所需物理量

第二階段將資料轉換為資料處理過程的固定格式，並且將缺失值(missing value)濾除。由於 ST 的原始資料格式與一般氣象通用的形式不同，例如原始速度資料中的風速的單位為公里每小時，而非氣象作業中慣用的單位，且有許多非氣象參數之儀器狀態資訊。故原始資料需要進行篩選並透過基本的單位轉換，以符合一般通用探空資料的格式。另外，在本階段中也同時計算校驗過程中所需使用的物理參考變數，例如：將風向風速轉換成水平風場(U,V)；利用氣壓、溫度、相對濕度換算比濕(q)；利用資料時間與空間位置資訊換算晴空輻射量等。這些資訊都將前處理過的資料使用，亦可直接輸入第三階段的校正模型進行校正。

### (三) 機器學習校正

第三階段利用本實驗室團隊以廣義線性模型 (Generalized linear model, GLM) 及梯度提升模式 (Gradient Boosting Machine, GBM) 作為核心，所發展出的溫度及濕度的校正模組 (Chang et al., 2021)，資料經過校正後的產出 L3 資料。在實際作業時我們經常會遇到部分資料缺失的情形，而校正流程中所使用之校正策略 (校正模組) 會因應資料的狀況自動進行辨識與判斷。舉例來說，在氣壓、溫度、相對濕度和晴空輻射量值皆無缺失時，將先以所有氣象參數作為輸入，得到 ST 溫度的修正量 ( $\Delta T$ )；再將這些氣象變數配合修正後的溫度作為輸入得到  $\Delta q$  作為比濕之修正量。最後透過後處理流程，將修正後的比濕、溫度換算為修正後的相對濕度並輸出。但是，如果原始資料中相對濕度有缺失但其天參數正常時，將利用另一組以除相對濕度外的參數及晴空輻射量值作為輸入所訓練出來的模型進行修正，也僅針對溫度進行誤差修正。

#### (四) 等垂直間距資料產製

在部分的氣象應用上，使用非等間距的探空資料對於資料比對或是銜接後續應用程式上會有一定的困擾。所以依照 Ciesielski et al. (2012) 的建議，我們在第四階段也將校正後的資料進行均一化處理，透過資料內插的方式，使資料維持在相同高度間距 (1hPa)。嚴格來說，在資料校正流程中，機器學習模型是以 ST 原始的垂直解析度資料進行校正，但為了後續資料相互比對及分析的便利性，可以透過固定間距為 1hPa 的垂直內插處理讓不同解析度資料之間能夠直接進行比較。

#### (五) NCAR-ASPEN 後處理

在很多的氣象研究過程中，我們都會使用 NCAR-ASPEN 作為探空資料後處理的步驟之一，而 ST 的校驗流程中我們僅針對缺失資料進行篩選與標注，並沒有對於資料的特性進行篩選。因此最後一個階段，我們也開發了銜接利用 ASPEN 軟體做後處理的模組，利用 ASPEN 內建的資料檢核機制，進行資料篩選與過濾。流程中，我們統一將每個階段產製的資料轉換為大氣領域通用的 EOL 格式。

### 三、結果

整體統計來說，以氣象局作業用儀器 Vaisala RS-41 做為比較基準，在溫度部分 (圖三左)，ST 作為原始觀測資料，隨著高度上升誤差越大，且 CDF

校正的方式並沒有很好的修正誤差，而 GBM 的部分修正誤差的表現最好，尤其在低層 (1000-700 hPa) 的資料誤差都有明顯的降低。在水氣量觀測部分 (圖三右)，雖然透過 GBM 校正過後誤差還是有隨高度上升的趨勢，但在低層 (1000-700 hPa) 的表現較其他方式更好。

而在個案當中，以 2020 年 11 月 20~24 號 YESR2020 觀測期間的探空時序列為例 (圖四)，可以看到 ST 在未校正前相較 Vaisala 的低層水氣有顯著高估的情形，這與原始資料中的溫度暖偏差有很大的關係。在經過 GBM 校正後有明顯改善，顯示此流程配合機器學習的校正，能有效將探空資料的誤差降低，並且流程能將資料格式固定，輸出對使用者較為友善的格式。

### 四、結論及未來工作

在本研究中，校正流程確實將溫度及水氣量方面的誤差縮小，且有不錯的效果。流程的其他優點包括：高度模組化且可攜性高，可以切換到不同系統平台上運作並且方便維護更新；流程效能佳，平均能在 10 秒內處理一份資料並產製校正後的檔案，在面對野外觀測環境下可以提供立即的分析使用；資料輸出為氣象常用的 EOL 格式，未來在面對不同軟體上得應用都可以滿足需求等等。

未來工作將比對不同機器模型的校正效果、和文化大學水文資料庫合作，將流程整合為線上網頁方便觀測人員使用並建檔、納入地面測站資料校正儀器誤差 (ground check system)、探空資料分析套件等等模組開發項目將陸續擴充，以完備 ST 資料校正系統。

### 五、參考文獻

- Nuret, M., Lafore, J., Guichard, F., Redelsperger, J., Bock, O., Agusti-Panareda, A., & N'Gamini, J., 2008, Correction of Humidity Bias for Vaisala RS80-A Sondes during the AMMA 2006 Observing Period. *J. Atmos. Ocean. Technol.*, **25**(11), 2152-2158.
- Ciesielski, P. E., Johnson, R. H., & Wang, J., 2009, Correction of Humidity Biases in Vaisala RS80-H

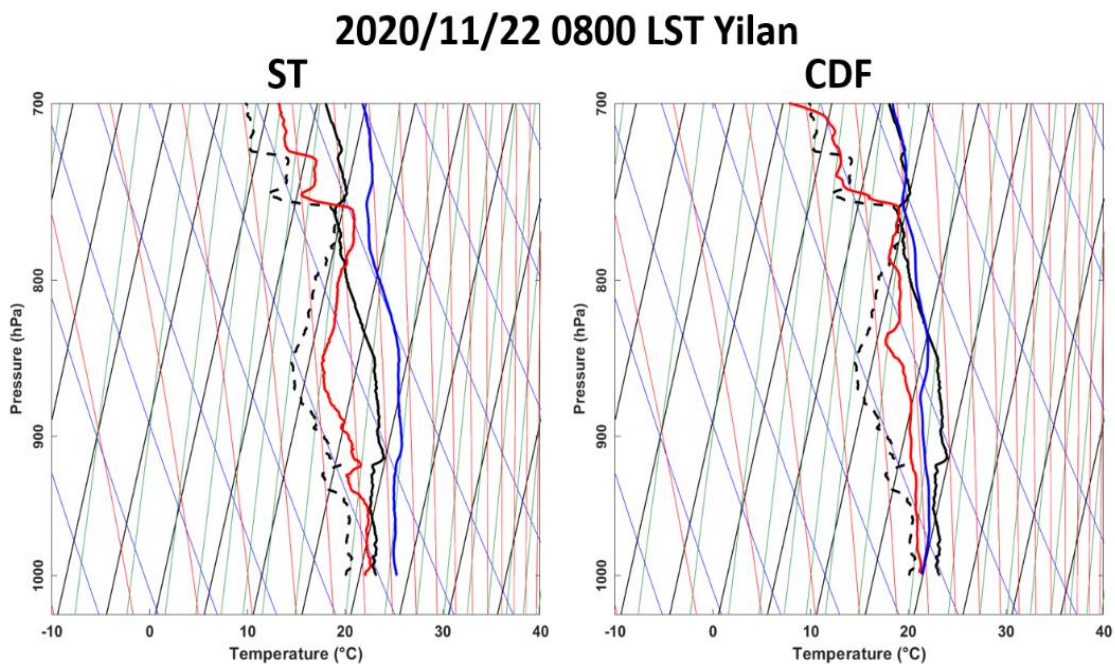
Sondes during NAME. *J. Atmos. Ocean. Technol.*, **26(9)**, 1763-1780.

Ciesielski, P. E., Haertel, P. T., Johnson, R. H., Wang, J., & Loehrer, S. M., 2012, Developing High-Quality Field Program Sounding Datasets. *Bull. Amer. Meteor.*, **93(3)**, 325-336.

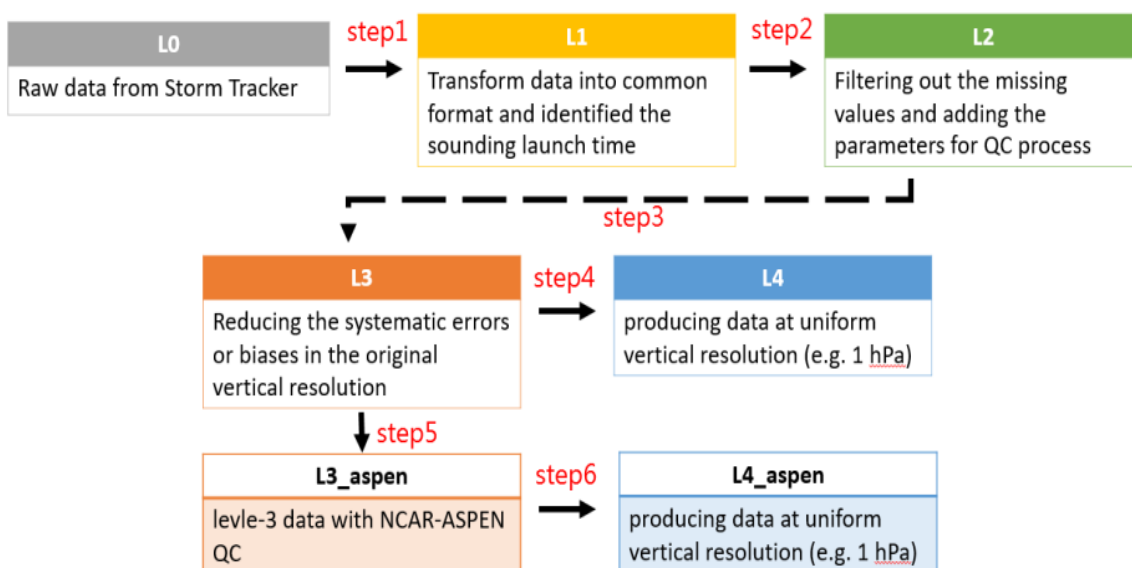
Hwang, W.-C., Lin, P.-H., and Yu, H., 2020, The development of the “Storm Tracker and its applications for atmospheric high-resolution upper-

air observations. *Atmos. Meas. Tech.*, **13**, 5395–5406.

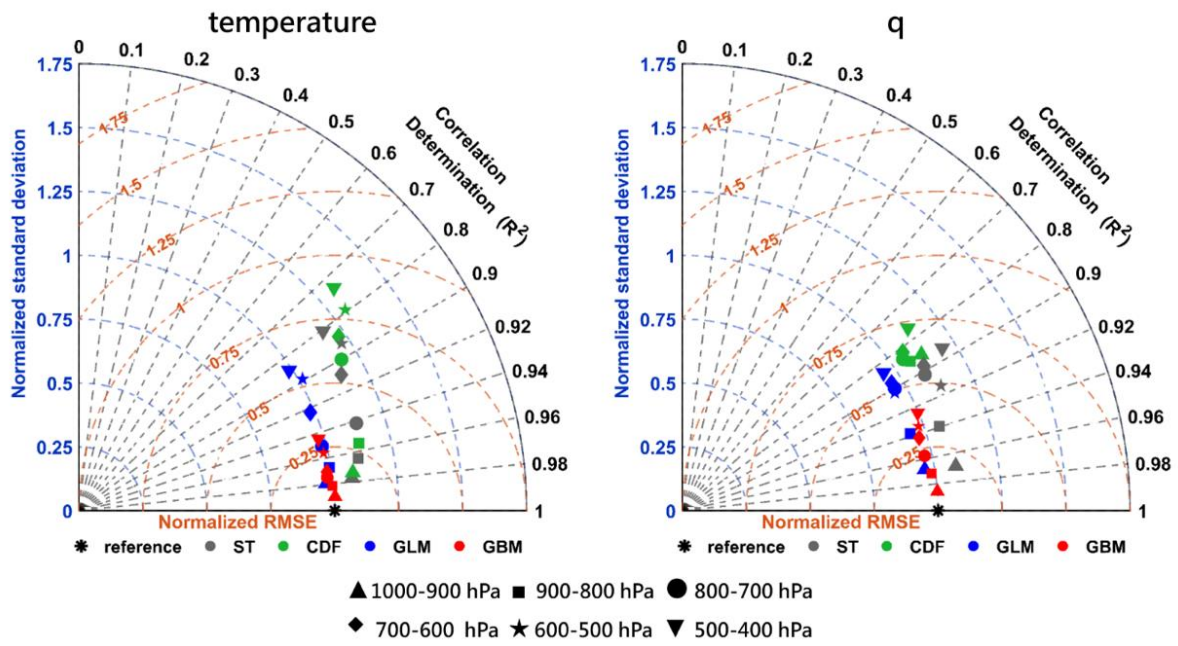
Chang, C.-W., Su, S.-H., Yu, T.-S., Wu, C.-M., Lin, P.-H., 2021, An artificial-intelligence-based approach of the mini-radiosonde-“Storm Tracker” data calibration. *Proceedings 2021 Conference on Weather Analysis and Forecasting*, A7-18. Central Weather Bureau, Ministry of Transportation and Communication.



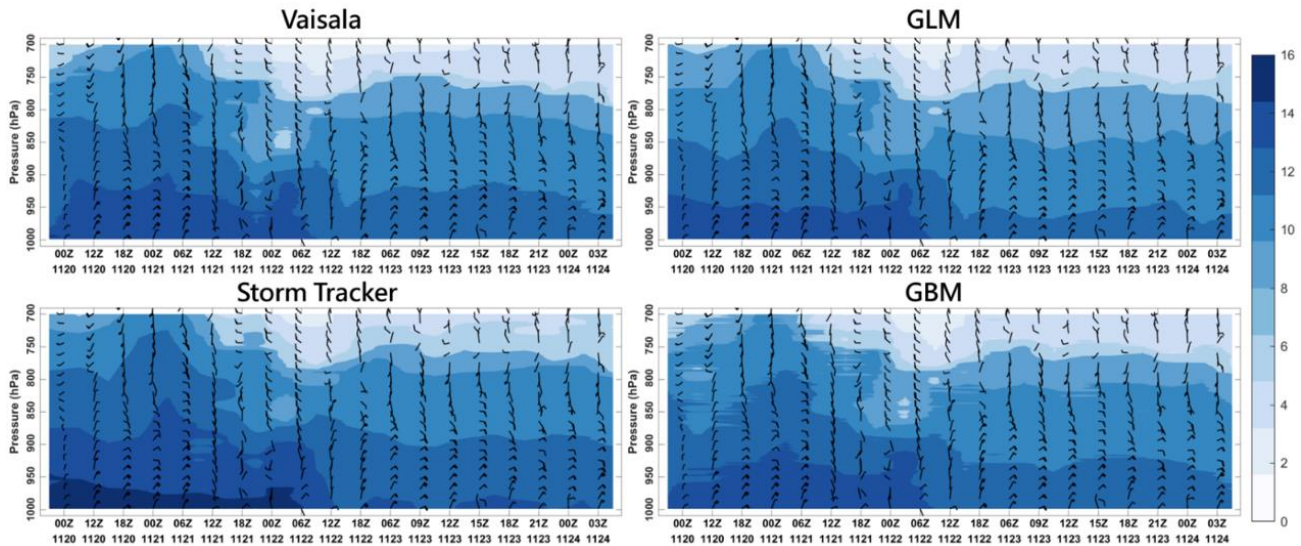
圖一、2020/11/22 00Z 之 Vaisala 與 ST、CDF 資料對比探空圖。黑色線為 Vaisala 資料，實線、藍線為溫度，紅線、虛線為露點溫度。



圖二、流程 workflow 簡圖



圖三、探空資料統計泰勒圖。藍色半徑為標準化的標準差，橘色同心半圓為標準化後的均方根誤差，黑色圓弧為相關係數。點的顏色表示不同資料種類。點的形狀代表不同高度區間。reference 代表以氣象局作業用儀器 Vaisala RS-41 為比較基準，越靠近基準表示誤差越小。



圖四、YESRain 期間探空時序列圖。著色部分為比濕 q