國立台灣大學共同教育中心統計碩士學位學程

碩士論文

Master Program in Statistics

Center for General Education

National Taiwan University

Master Thesis

離群值自動檢測系統應用於時雨量資料品管

An Automated Outlier Detection System for Hourly Rainfall Data Quality Control

周卉敏

Hway-Min Chou

指導教授：鄭克聲 教授

Advisor: Professor Ke-Sheng Cheng

中華民國 109 年 7 月

July 2020

doi:10.6342/NTU202001579

摘要

近十年來，資料品質保證越來越受到水文領域的重視。有了品質良好的降雨資料，才能確保使用它們進行水文應用相關的風險分析及決策管理時獲得可靠的研究結果。臺灣中央氣象局管理著一個由超過 600 個氣象站組成的自動雨量計網路系統，每日提供即時降雨觀測。有時一個雨量站觀測到的降雨量會明顯高於或低於附近其他測站的降雨量觀測值，由於相鄰測站的降雨量往往高度相關，這可能表示異常值存在於這些觀測值中。為了控制降雨資料的品質，我們必須將這些異常值區分出來。然而，目前為止，我們缺乏明確的標準以有效地判別。

在本研究中，我們運用統計方法以建立一個自動時雨量的異常值檢測系統。首先，我們根據臺灣四種常見的降雨類型的雨季，將收集到的時雨量資料分為四組。接著利用 K-Means 分群法對欲研究的雨量站按其地理位置和不同的降雨特性進行分群。然後，我們分別對每一種降雨類型的每一群進行主成分分析，計算出前幾個主成分，並建立一個表示降雨量資料異常程度的指標。

一旦某個測站的降雨量觀測值符合我們定義異常的指標，我們便可以立刻找出可能發生異常值的測站。最後，我們建立了自動離群值檢測系統，並將其呈現為線上的互動式網頁。本研究的目的在於對時雨量觀測值建立一個可靠的異常值檢測系統，使我們能有效地篩選出可能發生異常值的測站，以達到時雨量資料品質控制的目標。

關鍵字: 時雨量，資料品質控制，離群值檢測，主成分分析，K-Means 分群法

doi:10.6342/NTU202001579

Abstract

Data quality assurance has been receiving increasing attention in the field of hydrology in the last decade. Only high-quality data ensures data-driven risk analysis and decision-making strategies of hydrology applications. In Taiwan, the Central Weather Bureau manages an automated rain gauge network system of over 600 stations to obtain real-time precipitation observations. Occasionally, rainfall observations of one station are markedly higher or lower than those of nearby stations, suggesting the presence of anomalies because rainfall observations of neighboring stations are often highly correlated. To obtain reliable results based on hourly rainfall data, these anomalies should be identified in advance. However, there is a lack of definite criteria for effectively identifying anomalies.

In this study, we established an automated anomaly detection system for precipitation observations. First, we categorized the data into four groups according to the four fundamental storm types in Taiwan (frontal rain, Meiyu, convective storms, and typhoons). Second, we adopted K-means clustering analysis to classify all rain gauge stations of interest by their geographical location and rainfall characteristics. For each cluster, principal component analysis was conducted to acquire the first few principal components, aiming to construct an index representing the extent of anomalies. Once

doi:10.6342/NTU202001579

the criteria are determined, identifying anomalies is straightforward. Eventually, we established the detection system and presented it as an online interactive web page. Thus, in this study, a dependable anomaly detection system was created for effectively screening out possible anomalies to achieve hourly rainfall data quality control.

Keywords: Hourly Precipitation, Data Quality Control, Anomaly Detection, Principal Component Analysis, K-Means Clustering Analysis

doi:10.6342/NTU202001579

Table of Contents

doi:10.6342/NTU202001579

List of Figures

doi:10.6342/NTU202001579

List of Tables

doi:10.6342/NTU202001579

1. Introduction

Rainfall data are essential to agricultural farming, travel planning, and performing nearly all daily activities. The Central Weather Bureau (CWB) manages an automated rain gauge network system of over 600 stations to obtain real-time precipitation observations in Taiwan. Countless decisions required for livelihood activities rely on the analyses of these rainfall observations. Accordingly, the quality of rainfall data is paramount, necessitating rainfall data quality assurance (QA) and rainfall data quality control (QC). Data QA investigates inconsistencies and anomalies in the original data. Data QC uses the information from the QA process to determine whether the data can be used for analysis or applications. QA approaches utilized in manufacturing have wide applications, including observation, data archiving, and processing and dissemination of environmental information (Hudson et al., 1999). In the field of hydrology, data QA has been received increasing attention (You et al., 2007; Branisavljević et al., 2009).

Occasionally, anomalies occur the hourly observations provided by rain gauge stations. For example, when a station fails to send the observations in time because of malfunctions, delays, or unknown reasons, the amount of delayed observation becomes

doi:10.6342/NTU202001579

exceptionally high because it has been accumulating for several hours. Moreover, the rainfall data returned by a station may be notably higher or lower than those reported by nearby stations, suggesting the presence of anomalies because the rainfall amounts of neighboring stations are often highly correlated. To guarantee the reliability of hourly rainfall data, these anomalies must be identified. However, no definite criteria exist for instantly and effectively discovering these anomalies, and manual identification would be inefficient and infeasible. Therefore, in this study, we established an automated anomaly detection system for hourly precipitation observations. Using this system, rainfall data QC can be accomplished in a cost-effective manner.

Different methods have been reported for detecting rainfall anomalies. You et al. (2007) proposed three approaches for data QA of daily precipitation. First, to establish a QA test, they used a single gamma distribution with estimated statistical parameters instead of a normal distribution. Only values crossing a certain threshold are considered anomalies. Second, they developed the Q-test using a metric based on comparisons with neighboring stations. Third, they developed the multiple interval gamma distribution (MIGD) method, which assumes that meteorological conditions that produce average precipitation at surrounding stations will result in a predictable range at the target station.

doi:10.6342/NTU202001579

This method bins the average rainfall at nearby stations, and for events in a particular bin, it derives a gamma distribution by fitting the same events for the target station. Eventually, a QC test can be performed using the threshold of the new gamma distribution. The aforementioned three approaches consider the relationship between the rainfall at surrounding stations and the rainfall at the target station. However, they do not consider the rainfall characteristics of different seasons. Moreover, these approaches are more suitable when the number of stations required is relatively small.

Toe et al. (2017) conducted K-means cluster analysis and principal component analysis (PCA) to investigate the spatial and temporal variation patterns in the Central Dry Zone (CDZ) of Myanmar. They considered the influence of the climatological monsoon break on precipitation in the CDZ. Additionally, they divided the stations into different clusters to reveal the orographic effect and distinct climate dynamics. Their data revealed that the first and second principal components (PCs) mainly accounted for the spatial variabilities and seasonal (temporal) variation in average monthly precipitation in the CDZ, respectively. Before employing PCA, Toe et al. (2017) performed clustering to classify the original stations. Stations belonging to the same cluster possess

doi:10.6342/NTU202001579

similar rainfall characteristics. Furthermore, the obtained PCs could fully capture both spatial and temporal variations in precipitation.

In this study, we used statistical methods to generate the criteria for identifying anomalies. Because rainfall amounts are greatly affected by various rainfall characteristics (Boyle & Chen, 1987; Chen et al., 1999; Chen & Chen, 2003), we grouped the stations of interest to identify anomalies. Inspired by the method of Toe et al. (2017), we conducted adopt K-means cluster analysis (Cox, 1957; Fisher, 1958; Engelman & Hartigan, 1969) of the stations based on the features related to geographical locations and primary storm types in Taiwan (Wang & Cheng, 1982). Then, we performed PCA (Pearson, 1901; Hotelling, 1933; Jolliffe, 2002) for detecting outliers.

Rousseeuw and Hubert (2018) proposed the PCA outlier map for detecting outliers in a data set. They used three-dimensional data and fitted the data with two PCs. The map's vertical axis measures the orthogonal distance, which is the Euclidean distance of the data point to its 2-dimensional projection. The horizontal axis represents the score distance, which is the Mahalanobis distance of the data's projection relative to all

doi:10.6342/NTU202001579

projected data points. Both high orthogonal and score distances indicate a possible outlier. In the present study, we used a similar approach. We plotted the newly transformed coordinates of the first two PCs. Then, we defined that once the Euclidean distance between any projected data point and the origin of the plot exceeds the threshold (Section 3-4) we set, suggesting the existence of anomalies.

The rest of the paper is structured as follows. Chapter 2 presents data collection and preprocessing. Chapter 3 illustrates the two methods used to identify the two types of anomalies. The K-means clustering results, detected anomalies, and the system we developed are presented and discussed in Chapter 4. Chapter 5 provides the conclusion.

doi:10.6342/NTU202001579

2. Data

This chapter illustrates the data collection and analysis process. The CWB operates a rain gauge network of more than 600 stations around the country. We used the hourly rainfall data recorded by 297 rain gauge stations (Appendix A) set up by the CWB because they provide consistent rainfall data of better quality. The unit of each hourly rainfall is millimeter per hour.

Next, we web-scraped the hourly rainfall data from January 1, 1998, to May 30, 2020, from the Central Weather Bureau Observation Data Inquire System (CODiS). CODiS is an online open data platform that offers free observation data of CWB's automatic weather stations. Once users input the city where the preferred station is, the periodicity of data, and the period, CODiS will generate a report of weather data. We inspected the result page and ran the Python script to access and extract the column named Precp (24 observed precipitation a day) for each station.

We then preprocessed the collected data according to the rainfall characteristics of Taiwan. Taiwan is affected by the northeasterly monsoon from September to April and the southwesterly monsoon from May to August each year (Boyle & Chen, 1987; Chen et al., 1999; Chen & Chen, 2003). Wang and Cheng (1982) categorized the rainfall regimes in Taiwan into five categories:

doi:10.6342/NTU202001579

1. Winter (from December to February)
2. Spring transition (March and April)
3. Mei-yu season (from mid-May to mid-June)
4. Typhoon season (from mid-July to August)
5. Autumn rainfall (from September to November)

We simplified the classification of Wang and Cheng (1982) into four regimes (Table 1) representing the four main storm types of Taiwan: frontal rain, Meiyu, convective storms, and typhoons.

1. Frontal rain: the rainfall caused by the northeast monsoon and the spring rainfall, which are caused by the frontal systems to northern Taiwan.
2. Meiyu: a type of stationary front that usually occurs in May and June. It forms when the warm and cold fronts meet, and neither of them has the force to move the other.
3. Convective storms: the sun heats the ground, resulting in warm air rising, which cools to form heavy clouds. When rainstorms occur (often from July to October), they usually include thunder and lightning and have a short duration (the length of time that a rainfall event lasts at an observed location).

doi:10.6342/NTU202001579

4. Typhoons: a region-specific term for a tropical cyclone that usually occurs within the northwestern region of the Pacific Ocean and west of the International Date Line, with a much higher duration than convective storms.

The hourly rainfall data were divided into four groups according to the rainy seasons, as they have different rainfall characteristics. Table 1 presents the rainy seasons and duration of the four storm types. We easily separated frontal rain and Meiyu by their rainy seasons. However, both convective storms and typhoons tend to occur from July to October. To successfully distinguish the two events, we considered the duration of each rainfall event from July to October. Furthermore, we referred to the list of warning typhoons from 1998 to 2020 (Appendix B) issued by the CWB. If the duration of a rainfall event exceeded 12 h and corresponded with a typhoon warning, hourly rainfalls of that event were classified as typhoons instead of convective storms.

Table 1.

Rainy Seasons and Duration for Four Storm Types

Storm Type	Rainy Season	Duration
Frontal Rain	Nov. - Apr.	More than 1 hour
Meiyu	May and June	-
Convective Storms	July - Oct.	From 1 to 12 hours
Typhoons	July - Oct.	More than 12 hours

Note. Duration = The length of time that a rainfall event lasts at an observed location or in a particular area.

doi:10.6342/NTU202001579

3. Methods

Although anomalies have many causes, the effective identification of the anomalies is crucial. This chapter illustrates the methods for establishing the criteria for identifying anomalies. On examining the hourly rainfall data, we found two primarily abnormal situations. First, failure to return observations in time due to malfunctions or delays. Second, the time series of hourly rainfalls of a station differs markedly from that of nearby stations. We used the cutoff point method to identify the former situation and PCA to identify the latter situation. This chapter is structured as follows: Section 3-1 describes the two abnormal situations. Section 3-2 introduces the cutoff point method. Before adopting PCA, K-means clustering analysis was performed, the results of which are presented in Section 3-3, to group 297 stations for each storm type. Finally, Section 3-4 describes the PCA method.

doi:10.6342/NTU202001579

3-1. Abnormal Situations

Situation I: Failure to return observations because of malfunctions or delays.

The CWB defines the following codes in Table 2 for specific circumstances that occur when a rainfall gauge station returns observations. Codes-9991, -9995, -9997, and -9999 all denote different circumstances with no observations. Code-9996 indicates that an instrument delayed returning observations, causing the value of hourly rainfall to accumulate over a period before being returned. Code-9998 indicates that the observed amount of rain is minimal. We reorganized these codes into three categories—A1, A2, and A3.

Table 2.

Reorganization of CWB's Codes Corresponding to Specific Circumstance

New Code	Circumstance	CWB Code
A1	The instrument observed trace	-9998
A2	The instrument failed to return OBS due to malfunctions or unknown reasons	-9991 -9995 -9997 -9999
A3	The instrument delayed returning accumulated OBS for a while	-9996

Note.
Trace = An amount of precipitation that is smaller than 0.1 millimeter;
OBS = Observations;
-9991 = Instrument malfunctions waiting for repair;
-9995 = The instrument failed to return OBS due to malfunctions;
-9996 and A3 have the same meaning;
-9997 = The instrument failed to return OBS for unknown reasons;
-9998 and A1 have the same meaning;
-9999 = The instrument did not observe rainfall.

doi:10.6342/NTU202001579

Situation II: Rainfall time series of a station differs markedly from that of nearby stations.

When hourly rainfall data of a specific station are considerably lower or higher than those of neighboring stations, this may be an anomaly. Table 3 lists nine circumstances for a station in 1 day that may be abnormal and cause apparently different rainfall time series between a specific station and neighboring stations. B1 indicates that the recorded rainfall of the station at some hours is higher than that of nearby stations. B2 indicates that the observations are almost 0, indicated as “trace,” whereas neighboring stations have rainfall observations. B3 and B4 indicate a lack of recording rainfall because of mechanical failures (A2) and delays (A3), respectively, whereas nearby stations do. B5, B6, and B7 are the opposite of B2, B3, and B4, respectively. B8 represents that rainfall records have accumulated for a while before being returned. Finally, B9 indicates that the rainfall trend of the station is significantly different from that of nearby stations. B1, B2, B5, and B9 are circumstances requiring further verification for anomalies, whereas B3, B4, B6, B7 and B8 are obvious abnormal circumstances.

doi:10.6342/NTU202001579

Table 3.

Codes Corresponding to Specific Circumstances for a Station in One Day

Code	Circumstance	Need Verification
B1	OBS at some hours were higher than that of nearby stations	✓
B2	Observed trace; nearby stations, rainfalls	✓
B3	Did not observe OBS due to malfunctions	
B4	Did not observe OBS due to delays	
B5	Observed rainfalls; nearby stations, trace	✓
B6	Observed rainfalls; nearby stations did not due to malfunctions	
B7	Observed rainfalls; nearby stations did not due to delays	
B8	Delayed return of accumulated rainfall records	
B9	Rainfall trend was different from that of nearby stations	✓

Note.
Trace = an amount of precipitation that is $\leq$ 0.1 millimeter;
OBS = Observations.

doi:10.6342/NTU202001579

3-2. Cutoff Point

You et al. (2007) established the threshold approach for QC of daily precipitation of six specific stations, which fitted the daily observations to a gamma distribution for each station. This method is ideal when the number of stations to be analyzed is relatively small. However, it is not necessary to fit the daily precipitation to a gamma distribution for every station when the number of stations is large. Furthermore, our goal was to identify those accumulated observations that are exceptionally high (rainfall >150 mm/h) because of delayed returns. Therefore, we proposed the cutoff point method for detecting possible anomalies.

First, with the empirical distribution computed by hourly rainfall of every rain gauge station (0 values excluded), we obtained the cutoff point v_{1-p}. If the hourly rainfall $x(i,t)$ exceeds the threshold v_{1-p}, it was regarded as an anomaly.

$$x(i,t) \geq v_{1-p} \tag{1}$$

where p is a given probability, $x(i,t)$ denotes the hourly rainfall at station i and time t, and v_{1-p} denotes the $(1-p)^{\text{th}}$ quantile of the empirical distribution.

doi:10.6342/NTU202001579

3-3. K-Means Clustering Analysis

The second anomaly is the marked difference between rainfall data of a station from those of nearby stations. To effectively detect this type of anomaly, we employed the K-means clustering method to classify 297 rain gauge stations because rainfall characteristics vary with diverse geographical location and storm type.

According to Bock (2008), the K-means clustering approach is based on the sum-of-squares (SSQ) criterion. Several scientists in different fields under various assumptions have proposed different types of this K-means algorithm, and this method has been investigated and modified for decades. By either considering continuous analogs of the SSQ criterion (Cox, 1957; Fisher, 1958; Engelman & Hartigan, 1969) or studying the asymptotic behavior under random sampling strategies (Hartigan, 1975; Pollard, 1982; Bock, 1985), the application of the K-means algorithm has been extended to numerous novel data types and probabilistic models.

The K-means clustering method partitions a data set into K distinct and nonoverlapping clusters. Before clustering, the desired K clusters need to be determined. Then, the algorithm allocates each observation to one of the K clusters. Assuming n observations in our data set, $C_1, C_2, \dots, C_K$ denotes sets that include the indices of observations in each cluster. These sets satisfy the following two properties.

doi:10.6342/NTU202001579

1. $C_1 \cup C_2 \cup \ldots \cup C_K = \{1,2,\ldots,n\}$. Each observation belongs to at least one cluster.

2. $C_k \cap C_{k'} = \phi \quad, \forall \ k \neq k'$. Namely, the clusters are nonoverlapping.

The K-means clustering method aims to minimize the within-cluster variation among K clusters. The within-cluster variation of cluster C_k is denoted as $V(C_k)$, which yields the following equation:

$$\min_{C_1,\ldots,C_k} \{\textstyle\sum_{k=1}^{K} V\,(C_k)\} \tag{2}$$

Then, $V(C_k)$ is defined using the squared Euclidean distance.

$$V(C_k) = \frac{1}{|C_k|} \textstyle\sum_{x_i \in C_k} \parallel x_i - \bar{x}_k \parallel^2 \tag{3}$$

where $|C_k|$ denotes the number of observations in the k^{th} cluster, and $\bar{x}_k$ is the mean of cluster C_k (also called the cluster centroid).

$$\min_{C_1,\ldots,C_k} \{\textstyle\sum_{k=1}^{K} \frac{1}{|C_k|} \sum_{x_i \in C_k} \parallel x_i - \bar{x}_k \parallel^2\} \tag{4}$$

doi:10.6342/NTU202001579

The algorithm work to solves Equation (4) in the following way:

1. Each observation is randomly allocated a number from 1 to K, which serves as the initial cluster assignment.
2. Iterations occur until the alteration of assignments stops:

    (a) The centroid $\bar{x}_k$ is computed for each K cluster (i.e., the mean for the observations in cluster C_k).

    (b) Each observation is allocated to the cluster whose centroid is the closest, as defined by the Euclidean distance.

doi:10.6342/NTU202001579

3-4. PCA

K-means clustering on 297 rain gauge stations was performed by geographical location and storm type; it was found that the rainfall characteristics of the stations within each cluster were noticeably similar. Then, PCA was used to develop the criteria for the automatic system for detecting anomalies. In this section, we introduce PCA and describe how the standards were established.

K-means clustering on 297 rain gauge stations was performed by geographical location and storm type; it was found that the rainfall characteristics of the stations within each cluster were noticeably similar. Then, PCA was used to develop the criteria for the automatic system for detecting anomalies. In this section, we introduce PCA and describe how the standards were established.

PCA, a technique for summarizing the information of a data set, was developed by Pearson (1901), Hotelling (1933), and Jolliffe (2002); the PCA method developed by Jolliffe (2002) is the best modern reference. PCA reduces the dimensionality of multivariate data while preserving meaningful information as much as possible. It uses unsupervised learning, relying entirely on the input data itself instead of the corresponding target data. PCA transforms the original data to a new coordinate system. The new set of variables, known as PCs, is a linear transformation of the original

doi:10.6342/NTU202001579

variables. Each new variable is uncorrelated with other new variables. After projecting the initial data, the first coordinate lies in the direction with the largest variance, the second coordinate with the second largest variance, and so on.

The equation of PCA is given by

$$\boldsymbol{Z} = \boldsymbol{\Phi X} \tag{5}$$

where $\boldsymbol{Z}$ denotes the PCs, $\boldsymbol{\Phi}$ is a matrix of coefficients called loads determined by PCA, and $\boldsymbol{X}$ is a data matrix with n observations and a set of p features.

Equation (5) yields p linear transformations that form the PCs using the original variables. The first PC is written as

$$Z_1 = z_{i1} = \phi_{11}x_{i1} + \phi_{21}x_{i2} + \ldots + \phi_{p1}x_{ip},\ i=1, 2, \ldots, n \tag{6}$$

doi:10.6342/NTU202001579

This has the largest sample variance ($Var(Z1)$ is maximum) and is subject to the constraint that $\sum_{j=1}^{p} \phi_{j1}^2 = 1$. Without the constraint, these elements can result in an arbitrarily large variance. The remaining Z_i values are computed such that their variances are maximized and subject to another constraint, so that the covariance between Z_i and Z_j $(i \neq j)$ equals to 0. For example, the optimization problem is solved to obtain the first PC.

$$\max_{\phi_1,\cdots\phi_p} \sum_{i=1}^{n} z_{i1}^2 = \max_{\phi_1,\cdots\phi_p} \{\frac{1}{n}\sum_{i=1}^{1}(\sum_{j=1}^{p} \phi_{j1}\, x_{ij})^2\} \tag{7}$$

We calculated the matrix $\boldsymbol{\Phi}$ using the covariance matrix $\boldsymbol{S}$, which is written as follows:

$$s_{ij} = \frac{\sum_{k=1}^{n}(x_{ik}-\overline{x}_i)(x_{jk}-\overline{x}_j)}{n-1} \tag{8}$$

Therefore, the singular decomposition of $\boldsymbol{S}$ solves the PCA problem.

$$\boldsymbol{U}^T\boldsymbol{S}\boldsymbol{U} = \boldsymbol{L} \tag{9}$$

doi:10.6342/NTU202001579

where $\boldsymbol{L}$ is a diagonal matrix containing the eigenvalues of $\boldsymbol{S}$, and $\boldsymbol{U}$ is a matrix containing the eigenvectors of $\boldsymbol{S}$. $\boldsymbol{\Phi}$ can be computed by these two matrices.

$$\boldsymbol{\Phi} = \boldsymbol{U}\boldsymbol{L}^{-\frac{1}{2}} \tag{10}$$

If we scale the variables and make their variances equal to one, then $\boldsymbol{\Phi}$ is simply the eigenvector matrix $\boldsymbol{U}$. The covariance matrix becomes a correlation matrix $\boldsymbol{R}$.

$$r_{ij} = \frac{u_{ij}\sqrt{l_i}}{s_{jj}} \tag{11}$$

where u_{ii} is an element of $\boldsymbol{U}$. l_i is a diagonal element of $\boldsymbol{L}$ (λ_i) and s_{jj} is a diagonal element of $\boldsymbol{S}$. When $\boldsymbol{S}$ is replaced with $\boldsymbol{R}$, the principal components can be calculated by

$$\boldsymbol{Z} = \boldsymbol{\Phi}^T \boldsymbol{D}^{\frac{-1}{2}} \boldsymbol{X} \tag{12}$$

where $\boldsymbol{D}$ is the diagonal matrix obtained by $\boldsymbol{S}$ with each s_{jj} equals to one.

doi:10.6342/NTU202001579

Now, we establish the criteria for automatic detecting anomalies that may exist in the hourly rainfall observations. Our anomaly detection system analyzes the rainfalls within each cluster by different storm types in terms of spatial and temporal variables daily.

According to Toe et al. (2017), the first two principal components (PCs) obtained from the PCA fully explained the spatial and seasonal variations in the rainfall. Each PC is a linear combination of the normalized variables. PC loads represent the correlation coefficients of the normalized variables and a PC (Figure 3, Figure 5).

doi:10.6342/NTU202001579

I. PCA from Temporal Variation Aspect: to find the temporal variation in rainfalls at each station.

Given a specified cluster, the data matrix $\boldsymbol{X}$ of this cluster on one day is

$\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1m} \\ x_{21} & x_{22} & x_{23} & \cdots & x_{2m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & x_{n3} & \cdots & x_{nm} \end{bmatrix}$, where each column vector $X_j = \begin{bmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{nj} \end{bmatrix}$ denotes the hourly rainfall of n raingauge stations at hour j (the length of j must be at least larger than two). Next, we normalize each variable X_j to obtain the correlation matrix $\boldsymbol{R}$. The original data set is normalized because PCA computes a novel projection based on the standard deviation of the variables. A variable with an extremely high standard deviation will be given a higher weight for composing the new axis than a variable with a low standard deviation. If we normalize the data set in advance, then every variable will retain the same weight. By using Equation (12), we gain the first and second PCs Z_1 and Z_2.

Thereafter, we calculate the Euclidean distance between the origin and X_j being projected on the PCA subspace of the first two PCs.

$$d = \sqrt{Z_1^2 + Z_2^2} \geq d_{i,j,p} \tag{13}$$

Considering n as the number of days in which PCA can be performed, the number of rainy days with i^{th} storm type and j^{th} cluster is $n_{i,j}$. For these $n_{i,j}$ days, each day

doi:10.6342/NTU202001579

the maximum distance from the farthest projected data point to the origin can be computed. We obtained $d_{i,j,p}$ by taking the p^{th} quantile of those maximum $n_{i,j}$ distances and set $d_{i,j,p}$ as the threshold for determining anomalies. If d (the Euclidean distance from any projected data point to origin) exceeds $d_{i,j,p}$, this suggests the existence of anomalies at a specific station because PC1 (Z_1) captures the largest spatial variation, and PC2 (Z_2) accounts for the remaining variation of those normalized variables. The spatial variation explained by each PC is nonoverlapping.

II. PCA from Spatial Variation Aspect: determining the spatial variation in rainfalls at each hour.

Given a specified cluster, the data matrix $\boldsymbol{X'}$ of this cluster on one day is the transpose is $\boldsymbol{X}^{\mathrm{T}} = \begin{bmatrix} x_{11} & x_{21} & x_{31} & \cdots & x_{n1} \\ x_{12} & x_{22} & x_{32} & \cdots & x_{n2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{1m} & x_{2m} & x_{3m} & \cdots & x_{nm} \end{bmatrix}$, where each column vector $X_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{im} \end{bmatrix}$ denotes the m hourly rainfall at station i (the length of *i* must be at least larger than two). Next, each variable X_i is normalized to obtain the correlation matrix $\boldsymbol{R'}$. Similarly, the first two PCs are gained using Equation (12). Finally, d' is acquired using Equation (13) and compared with the threshold $d_{i,j,p}{}'$. If d' exceeds $d_{i,j,p}{}'$, this suggests the existence of anomalies at a specific hour because PC1(Z_1) captures the largest temporal variation, and PC2 (Z_2) accounts for the remaining variation of those normalized variables. The temporal variation explained by each PC is nonoverlapping.

doi:10.6342/NTU202001579

4. Results

In this chapter, the results are presented and discussed as follows. Section 4-1 contains the anomalies caused by malfunctions or delays detected using the cutoff point method. Section 4-2 describes the K-means clustering analysis of 297 stations based on features such as geographical locations and rainfall characteristics by the four storm types. Finally, Section 4-3 displays the possible anomalies identified using PCA.

4-1. Anomalies Discovered Using the Cutoff Point Method

We calculated the cutoff point with $p = 0.001$ (the rainfall value for the 99.9% corresponding percentile with empirical distribution) for each rain gauge station using rainfall observations from 1998 to 2019. Next, we attained 5106 hourly precipitations that exceeded the cutoff points of 297 stations. We gained the previous hourly rainfall for each detected value. Using Table 2, the observations were assigned the values of A2 and A3. Eventually, 205 values were labeled A3 and 15 as A2 (Tables 4 and 5). Table 4 presents A3 data only for hourly rainfall > 200 mm (see Appendix C for complete results).

doi:10.6342/NTU202001579

Table 4.

Identified Hourly Rainfall Anomalies Caused by Delay Returns

Item	Station ID	Station	Time	HR
1	C0A931	Sanhe	2008/11/09 05:00	936
2	C0A940	Jinshan	2008/12/24 19:00	735.5
3	C0AI40	Shipai	2008/10/21 14:00	283
4	C0C490	Bade	1998/07/14 15:00	388.5
5	C0M640	Zhongpu	2001/09/19 10:00	741.5
6	C0O970	Hutoupi	2004/10/21 13:00	201
7	C0R130	Ali	2001/05/21 11:00	544
8	C0R140	Majia	2001/05/21 11:00	828
9	C0R140	Majia	2001/05/31 10:00	434
10	C0R280	Binlang	2012/08/28 15:00	666
11	C0R341	Mudan	2011/09/03 13:00	320.5
12	C0R341	Mudan	2012/08/25 04:00	460.5
13	C0S660	Siama	2016/07/09 14:00	369
14	C0S710	Luye	2016/07/09 15:00	226.5
15	C0S760	Hongshih	1999/09/04 19:00	216.5
16	C0S760	Hongshih	1999/10/09 11:00	237
17	C0V310	Meinong	2001/05/21 10:00	312.5
18	C0V350	Xipu	2001/05/21 10:00	342
19	C0V740	Qishan	2001/05/21 11:00	292
20	C1A630	Siapen	2001/07/21 14:00	216
21	C1E480	Fongmei	1998/02/25 18:00	268
22	C1F891	Shaolai	1998/02/20 17:00	223.5
23	C1F941	Xueling	1998/02/20 18:00	251.5
24	C1R110	Gusia	2001/05/21 14:00	579.5
25	C1R110	Gusia	2001/05/31 16:00	334
26	C1R120	Shangdewun	2001/05/21 11:00	711
27	C1S670	Motian	2016/07/09 13:00	216.5
28	C1U690	Sinliao	2009/10/12 14:00	734.5
29	C1Z130	Tongmen	2005/09/23 09:00	364.5

Note.
HR = Hourly Rainfall (mm);
The instrument failed to return the observation in time; the value had accumulated for several hours.

doi:10.6342/NTU202001579

Table 5.

Identified Hourly Rainfall Anomalies Caused by Malfunctions

Item	Station ID	Station	Time	HP
1	C0A870	Wujhihshan	2000/08/18 16:00	75
2	C0D360	Meihua	2019/07/01 14:00	108.5
3	C0R150	Sandimen	2000/11/01 12:00	83
4	C0R190	Chishan	2000/11/01 14:00	85
5	C0R190	Chishan	2000/09/24 09:00	77.5
6	C0S760	Hongshih	2000/07/03 08:00	63
7	C0S760	Hongshih	2000/07/06 08:00	44.5
8	C0S760	Hongshih	2000/07/18 07:00	47.5
9	C0S760	Hongshih	2000/08/04 08:00	43
10	C0S760	Hongshih	2000/08/07 09:00	56.5
11	C0V350	Xipu	2000/04/18 16:00	82.5
12	C1H9B1	Amei	2019/06/14 16:00	131.5
13	C1I121	Da-An	2000/03/21 15:00	150
14	C1I121	Da-An	2000/11/09 14:00	158
15	C1I131	Tongtou	2000/11/09 14:00	103

Note.
HP = Hourly Precipitation;
The instrument failed to return the observation in time due to malfunctions.

Even though not every detected value was an anomaly, from Tables 4 and 5, we could successfully identify the abnormal observations (-9996 and -9995) through the cutoff point method.

doi:10.6342/NTU202001579

4-2. Results of the K-Means Clustering Method

Because the rainfall amount changes with the location and type of storm, we calculated the following five variables for each storm type and used them to conduct the K-means clustering analysis of 297 stations.

1. ALT: the altitude of a station.
2. $LONG$: the longitude of a station.
3. LAT: the latitude of a station.
4. $\bar{X}$: the average annual rainfall from 1998 to 2019 of a specified storm type.
5. sd: the standard deviation of the average annual rainfall from 1998 to 2019 of a specified storm type.

The first three variables are related to locations, whereas the other two variables captured the rainfall characteristics of each storm type. We examined the clustering outcomes for the number of clusters K= 4, 6, 8, 9, 10, 12, and 15 for the types of frontal rain, Meiyu, convective storms, and typhoons. Finally, we divided 297 stations into 10 clusters for convective storms and eight clusters for each of the other three storm types. The ideal clustering results of four storm types are presented in Figure 1 (See Appendix A for the detailed clustering result of each storm type).

doi:10.6342/NTU202001579

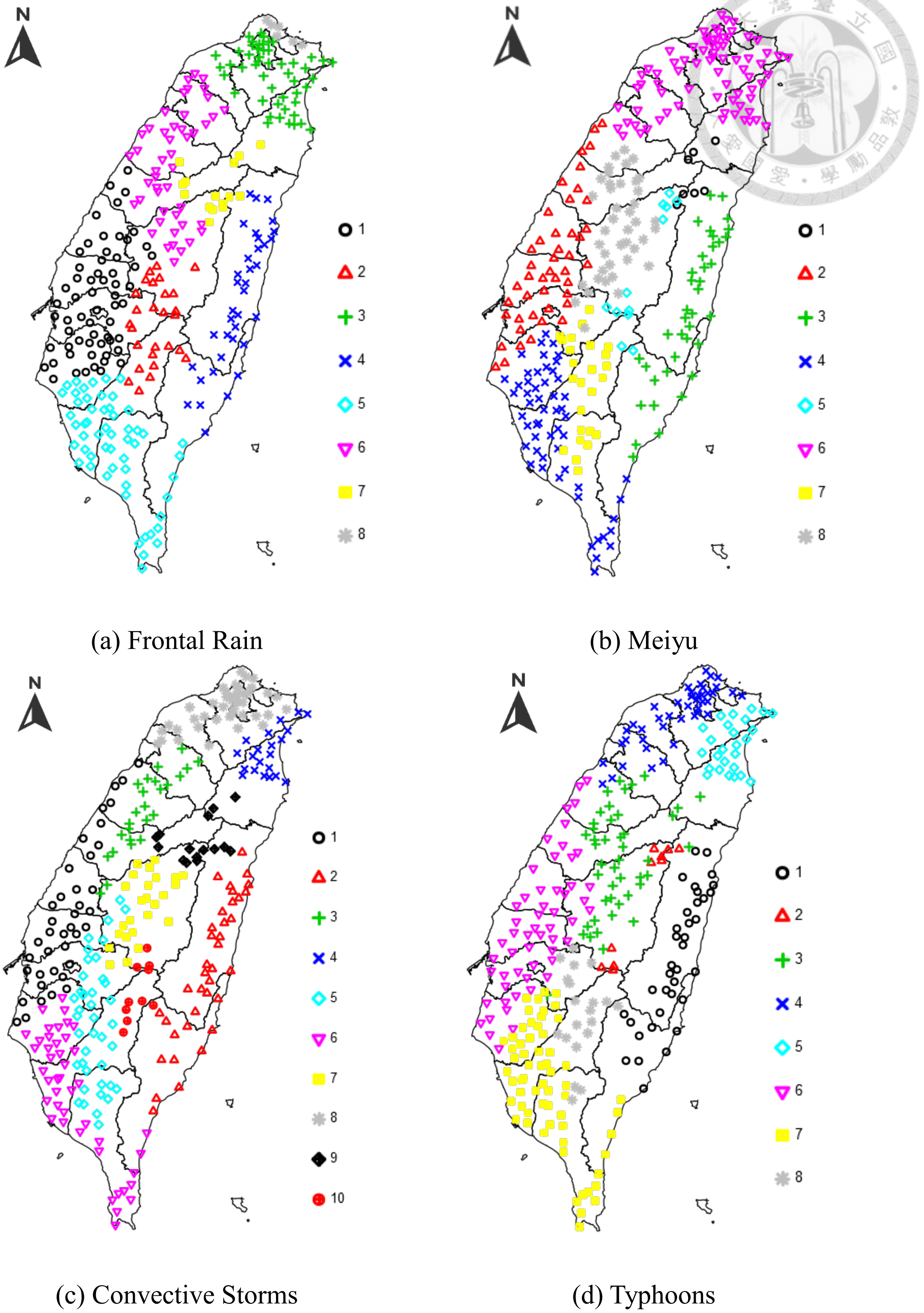
Taiwan is affected by the northeast monsoon from November to April. The monsoon strengthens and causes strong winds in the coastal areas and northern Taiwan. Some stations located in the northeast area receive rainfall in winter. From Figure 1-(a), we noticed that these stations were categorized into the same cluster (Group 8).

Both Siberian High and Pacific High affect Taiwan during the Meiyu period (in May and June). The prevailing wind at this time is usually from the southwest. Figure 1-(b) displays three clusters in the southwest area and only one cluster in the north and east areas.

A convective storm usually occurs on a summer (July to October) afternoon with heavy rainfall and for a short duration. When the sun heats the surface of the ground, the high temperature causes water to evaporate, cool, condense, and form tiny drops of water as it rises in the atmosphere. This process continues until rainfall occurs. Figure 1-(c) shows 10 clusters, because a convective storm often covers a small area.

On average, at least three or four typhoons hit Taiwan every year, primarily from July to October. Figure 2 presents the nine main paths of typhoons hitting Taiwan from the Typhoon Database established by the CWB. The clustering result for typhoons (Figure 1-(d)) seems roughly in line with those paths (Figure 2).

doi:10.6342/NTU202001579



(a) Frontal Rain (b) Meiyu

(c) Convective Storms (d) Typhoons

Figure 1. Clustering Results for Four Preliminary Storm Types

doi:10.6342/NTU202001579

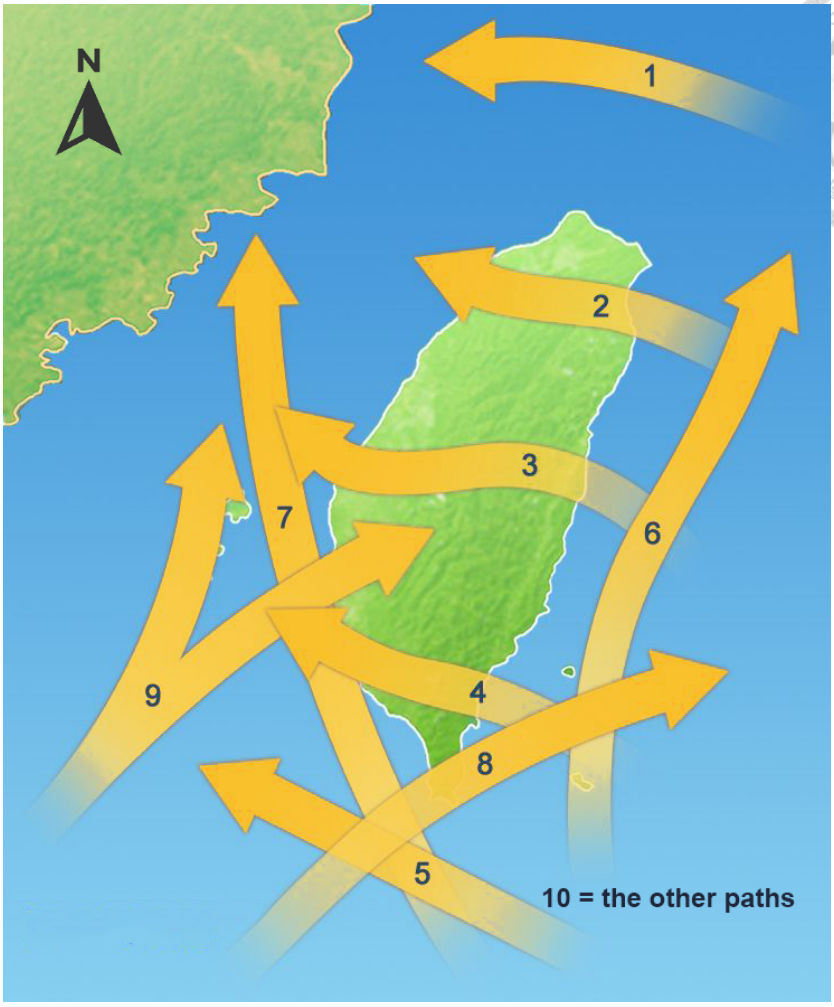


Figure 2. Nine Main Paths of Typhoons Hitting Taiwan
These paths are from the Typhoon Database of the CWB.

doi:10.6342/NTU202001579

4-3. Possible Anomalies Identified Using PCA

Following the clustering results, we performed PCA to establish the criteria for automated detection of anomalies for each storm type using the precipitation data from 1998 to 2019. We present the anomalies identified in this section. Section 4-3-1 demonstrates how to compute the Euclidean distance of the new transformed coordinate system, with one detected anomaly of the Donghe station on March 27, 2020, taken as an example. Section 4-3-2 describes the division of the identified anomalies of four storm types into nine categories (Table 3). Finally, Section 4-3-3 introduces the automated anomaly detection system.

doi:10.6342/NTU202001579

4-3-1. Detected Anomaly at Station Donghe on March 27, 2020

We employed the PCA method on the hourly precipitation of each day from 1998 to 2019. In short, the dimension of our data matrix was $n \times 24$ (data from 24 h a day of a cluster containing n stations). However, not all 24-h data were available for PCA for some days because it did not rain on those days. Therefore, we set two conditions that needed to be satisfied before performing the anomaly detection technique.

1. If the maximum hourly rainfall among all stations in the cluster is ≥ 5 mm, then retain the corresponding hour instead of removing it.
2. Retain the data of a day if it rained for at least 3 h that day; else, remove it

Condition 2 was included to reduce the dimensions from at most 24 to 2 using PCA, and it is unnecessary to use PCA if data are available from only 2 h (i.e., only two variables).

doi:10.6342/NTU202001579

I. Temporal Variation Aspect

Performing PCA from a temporal variation aspect enables us to observe the temporal variation patterns in rainfall for each rain gauge station. We took March 27, 2020, as an example. The data matrix of this day is a 36 × 9 matrix because Cluster 4 of the frontal rain type contains 36 stations (Table 7), and the 13th, 14th, 15th, 16th, 17th, 18th, 22th, 23th, and 24th h of this day satisfied the first condition. After normalizing the data matrix and conducting PCA, we obtained the variable correlation plot (Figure 3) and the new coordinate system (Figure 4) formed by PC1 and PC2.

First, we interpreted Figure 3. The horizontal axis represents PC1, which accounts for 50.5% variation in our original data matrix; the vertical axis represents PC2, accounting for 18.5% variation. Thus, the first two PCs explain 69% variation of the rainfall of this day. Figure 3 shows the correlation coefficients r between 2 PCs and the nine original variables (the 13th, 14th, 15th, 16th, 17th, 18th, 22th, 23th, and 24th h), which can be obtained as

$$r = \frac{v_{ij} \times e_j}{Std(X_i)} \tag{14}$$

doi:10.6342/NTU202001579

where v_{ij} denotes the i^{th} element of the j^{th} unit-length eigenvector of the covariance matrix, e_j denotes the eigenvalue of PC_j $(Var(PC_j))$, and $Std(X_i)$ denotes the standard deviation of the variable X_i. Because the data matrix is normalized, the value of $Std(X_i)$ is 1. Using Equation (14), the relationship between each PC and a specific variable can be obtained. For instance, the correlation coefficient of PC1 with the 13th hour is 0.84, whereas that of PC2 and the 13th hour is 0.12.

The colors in Figure 3 represent the expected contribution of a variable to the PCs. The contribution of a variable to a given PC (in percentage) is computed as follows:

$$contrib = \frac{r_{ij}^2 \times 100}{\sum_j \sum_i r_{ij}^2} \tag{15}$$

where r_{ij} denotes the correlation coefficient of variables X_i and PC_j. The expected contribution is attained using

$$\frac{\sum_j (contrib \times e_j)}{\sum_j e_j} \tag{16}$$

doi:10.6342/NTU202001579

where e_j denotes the j^{th} eigenvalue (variance) of PC_j. For example, the contributions of the 13th hour to PC1 and PC2 are 15.51% and 0.86%, respectively, whereas the expected contribution is approximately 11.57%.

Figure 4 displays the new coordinate system after the transformation. Similarly, the horizontal axis and the vertical axis of Figure 4 are PC1 and PC2, respectively. The dimensions are reduced from nine (hours) to two (PCs) for the precipitation data of 36 stations. Because PC1 and PC2 lie in the two directions with the first two greatest variances, the point that is the farthest from the origin indicates that the rainfall pattern of this station is much more distinct from that of the other stations. For each day available for PCA, we computed the Euclidean distance of each point to the origin and considered the largest distance. Given a cluster of a specified storm type, all these distances are obtained, and the threshold is determined using Equation (13) by setting p = 90th quantile. Thus, the criterion for detecting the anomalies is 10.185 (Table 7). The Donghe station is considered to have anomalies because its distance from the origin is 12.17, which exceeds the threshold of 10.185 (Figure 4). The other stations are very close to the origin except for the Hualien station (the distance = 5.63). This plot implies that the variation of rainfall of the Donghe station is mainly captured by PC1, whereas

doi:10.6342/NTU202001579

that of the Hualien station is explained by PC2. In other words, among 36 stations, the temporal variation of the Donghe station is the largest.

The colors in Figure 4 indicate the quality of representation of individuals. $\cos 2$ equals to squared r in Equation (14). A high $\cos 2$ indicates a good representation of the individual by the PCs, and a low $\cos 2$ means that the individual is not perfectly represented by the PCs. From the color of the point Donghe station, we find that it is well represented by PC1. Moreover, the Hualien station is represented by PC2.

doi:10.6342/NTU202001579

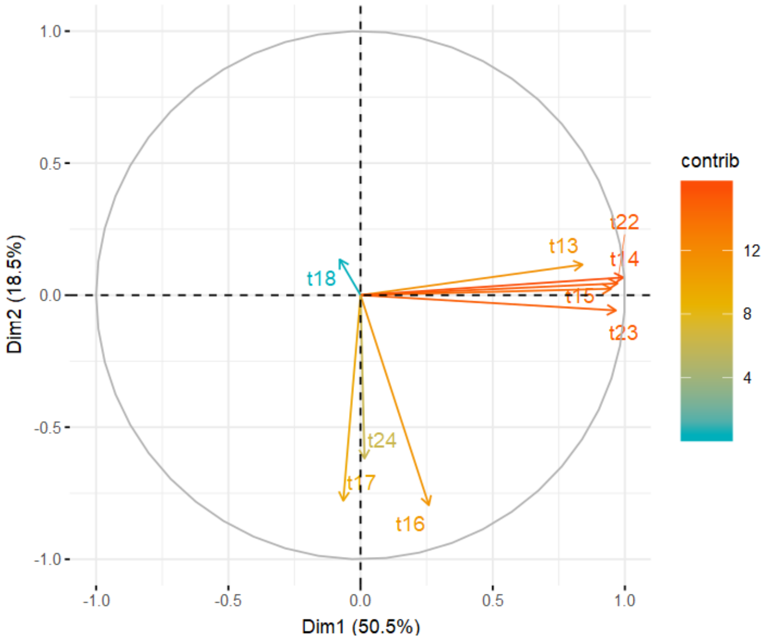


Figure 3. Variable Correlation Plot from Temporal Variation Aspect

The horizontal axis represents PC1; the vertical axis represents PC2. The colors represent the contribution (in percentage) of a variable to the principal components.

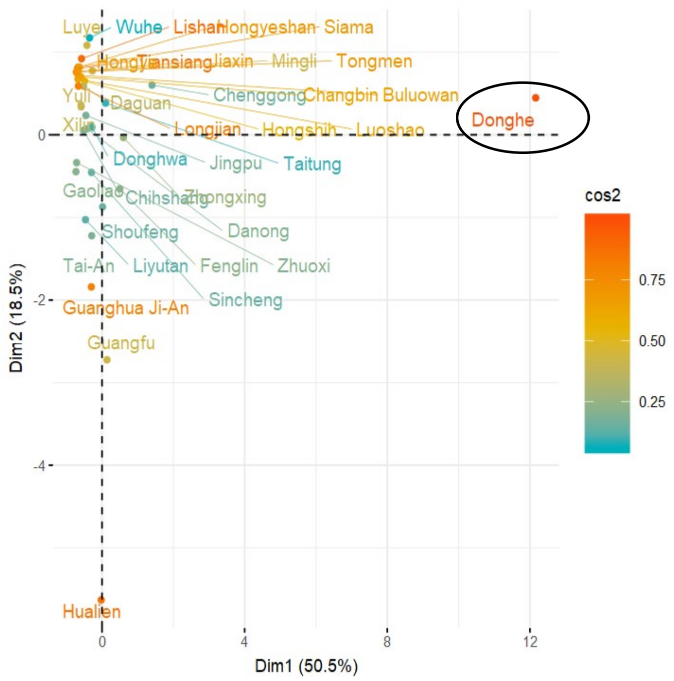


Figure 4. The New Coordinate System after PCA from Temporal Variation Aspect

The colors indicate the quality of representation of the individuals. Similar individuals are grouped together.

doi:10.6342/NTU202001579

II. Spatial Variation Aspect

Performing PCA from a spatial variation aspect helps us observe the spatial variation patterns in rainfall for each hour. The data matrix of this day is a 9 × 36 matrix. Each variable contains the nine hourly rainfalls: 13^{th}, 14^{th}, 15^{th}, 16^{th}, 17^{th}, 18^{th}, 22^{th}, 23^{th}, and 24^{th} h, which satisfies the first condition. Cluster 4 of the frontal rain storm type has 36 stations (Table 7). After normalizing the data matrix and conducting PCA, the variable correlation plot is obtained (Figure 5), and the new coordinate system (Figure 6) formed by PC1 and PC2 is also obtained.

In Figure 5, the horizontal axis represents PC1, which accounts for 33.7% variation in our original data matrix, and the vertical axis represents PC2, accounting for the other 21.8% variation, amounting to a total of 55.5% variation of the rainfall on that day. The circled stations in Figure 5 are the Donghe station and its seven nearby stations: in the order of distance, Chenggong, Chihshang, Luye, Hongshih, Mingli, Taitung, and Hongyeshan. The Donghe, Chihshang, Hongshih, and Mingli stations are negatively related to both PC1 and PC2, whereas the Chenggong, Luye, Taitung, and Hongyeshan stations are negatively related to PC1 and positively to PC2.

doi:10.6342/NTU202001579

Figure 6 displays the new coordinate system after the transformation. The horizontal axis and the vertical axis are PC1 and PC2, respectively. The dimensions are reduced from 36 (stations) to two (PCs) for the precipitation data of 36 stations. Compared with Figure 4, identification of the existence of anomalies is relatively harder in Figure 6. However, both the 15^{th} and 17^{th} h represent the 2 PCs well.

doi:10.6342/NTU202001579

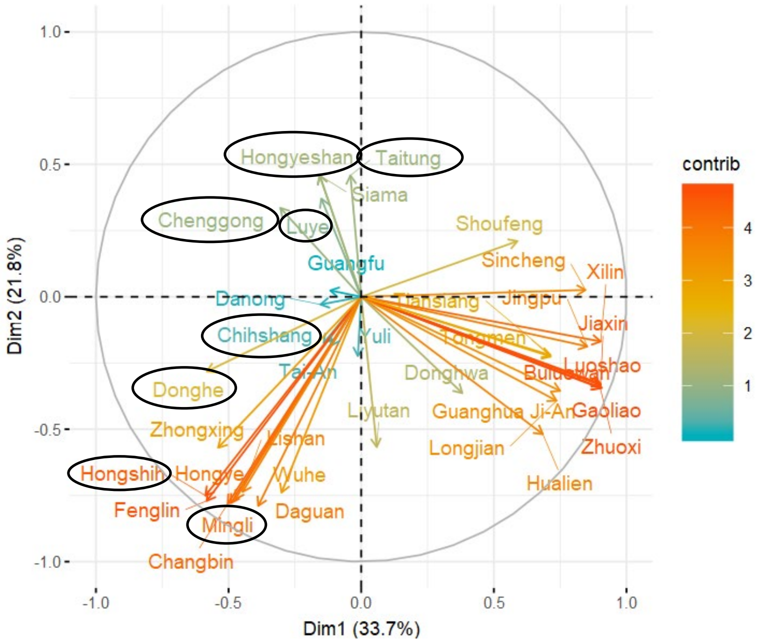


Figure 5. Variable Correlation Plot from Spatial Variation Aspect

The horizontal axis represents PC1; the vertical axis represents PC2. The colors represent the contribution (in percentage) of a variable to the principal components.

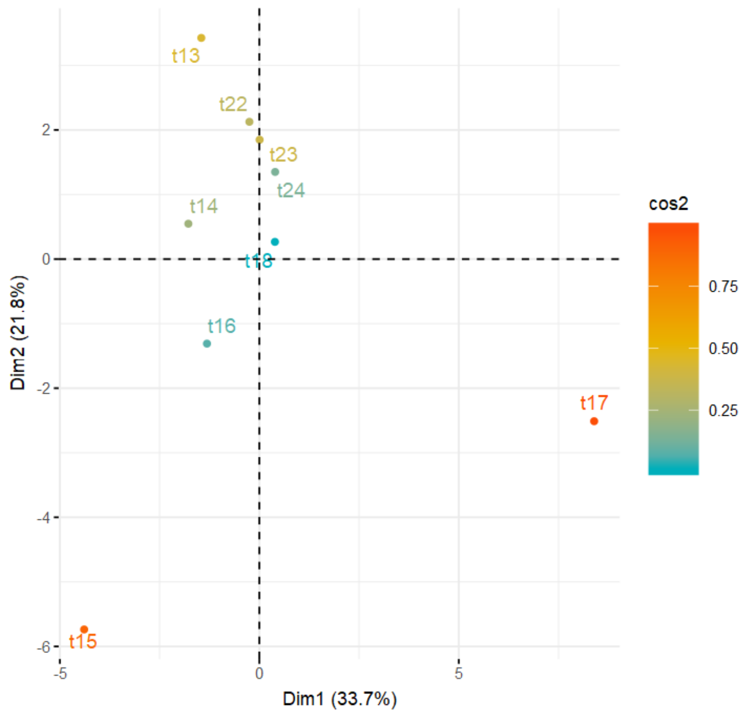


Figure 6. The New Coordinate System after PCA from Spatial Variation Aspect

The colors indicate the quality of representation of the individuals. Similar individuals are grouped together.

doi:10.6342/NTU202001579

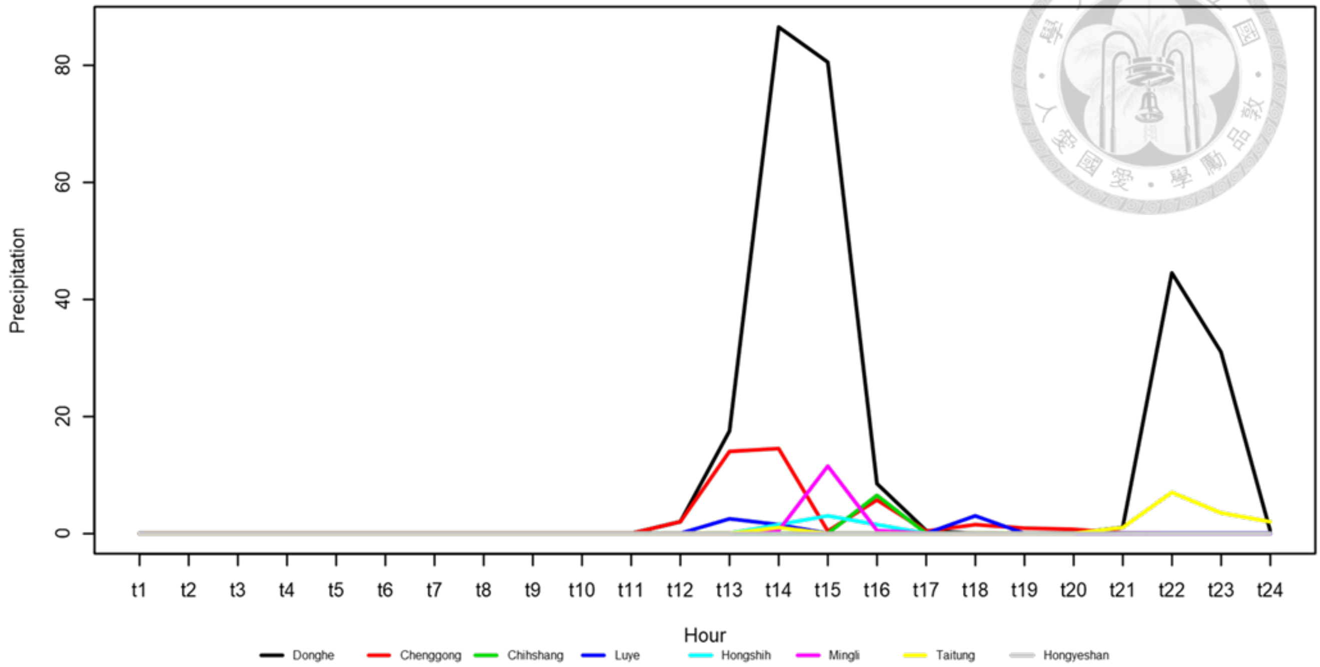


Figure 7. Rainfalls Observed by Donghe Station and Nearby Stations on Mar 27, 2020

Donghe Station is detected to have anomalies. The neighboring stations are Chenggong, Chihshang, Luye, Hongshih, Mingli, Taitung, and Hongyeshan, from near to far.

Figure 7 shows the hourly rainfalls of the Donghe station and the other seven stations. It rained a lot at the Donghe station from 1 pm to 4 pm and from 10 pm to 12 am on this day (recorded rainfall: 86.5 mm at 2 pm and 80.5 mm at 3 pm). Hence, the temporal variation pattern in rainfall of the Donghe station is quite different from that of the other neighboring stations, classified as B1 (Table 3). Although our system identified that the Donghe station might have anomalies, further verification is needed to ensure whether anomalies exist. Figure 8 shows the detected result on March, 27, 2020. The red cross

doi:10.6342/NTU202001579

represents the Donghe station, and the blue circles represents the other stations in Cluster 4 of frontal rain.

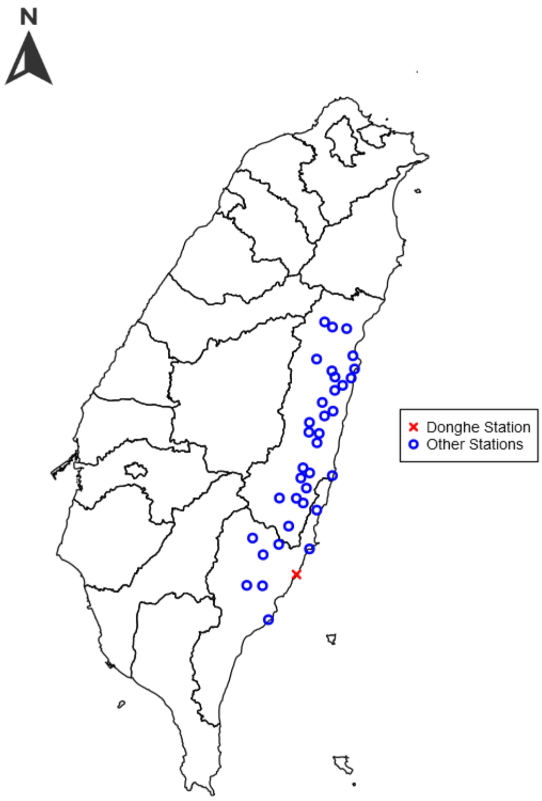


Figure 8. Anomaly Detected on Mar 27, 2020, in Cluster 4 of Frontal Rain

The red cross represents where Donghe Station is located, while there exist anomalies in the rainfalls that Donghe Station observed. The blue circles are other rain gauge stations, observing no anomalies, in Cluster 4.

doi:10.6342/NTU202001579

4-3-2. Criteria for Anomaly Detection and Nine Categories of Anomalies

Tables 6 presents the anomaly detection results of each cluster of each storm type. The available days for PCA of shows the number of days that satisfy the two conditions mentioned in Section 4-3-1. For each day, we computed the Euclidean distance from the subspaces of PC1 and PC2. Then, we obtained the 90th quantile distance as the criterion for detecting anomalies. Taking Cluster 1 of the frontal rain type as an example, we calculated 424 maximum distances and set 8.925 (90th quantile of these distances) as the threshold. The detected anomalies of each table present the number of anomalies (approximately one over ten of the available days) for each cluster of a specific storm type. After our system discovered these anomalies, we examined them thoroughly and identified the possible anomalies for each cluster by visual verification (PAIVV).

According to Table 3, PAIVV of all storm types were divided into nine categories and are presented in Table 7. Categories B1, B2, B5, and B9 require satellite or weather radar images for anomaly verification. By contrast, anomalies belonging to categories B3, B4, B6, B7, and B8 were successfully identified. Figure 9 shows that for each storm type, the number of category B1 is the most, and the number of B5 is the second most.

doi:10.6342/NTU202001579

Table 6. *Threshold for Anomaly Detection and Anomalies Detected of Four Storm Types*

Storm Type	Cluster	Number of Stations	Days Available	Threshold	Anomalies Detected	PAIVV
Frontal Rain	1	61	424	8.925	43	4
	2	27	514	6.859	52	7
	3	50	1401	12.039	140	23
	4	36	947	10.185	95	12
	5	55	551	10.391	55	11
	6	47	682	7.641	69	9
	7	15	592	6.093	60	5
	8	6	1153	5.234	116	5
Meiyu	1	7	392	4.511	40	8
	2	45	604	9.444	61	8
	3	38	618	9.239	62	17
	4	55	707	10.463	71	13
	5	12	653	5.906	66	9
	6	74	805	11.967	81	17
	7	25	774	6.947	78	12
	8	41	887	8.836	89	17
Convective Storms	1	39	787	8.61	79	6
	2	37	1084	8.708	109	13
	3	25	895	6.854	90	11
	4	21	929	7.841	93	9
	5	40	1568	8.715	157	18
	6	44	1363	9.952	137	7
	7	25	1077	7.703	108	16
	8	41	970	10.745	97	10
	9	15	843	6.228	85	3
	10	10	892	4.883	90	8
Typhoons	1	36	276	10.961	28	3
	2	10	308	7.484	23	1
	3	45	278	17.227	31	0
	4	43	285	13.033	28	3
	5	26	237	9.653	29	8
	6	55	237	10.343	24	2
	7	62	324	13.88	33	4
	8	20	280	10.626	28	1

Note. PAIVV = Possible anomalies identified by visual verification from the detected anomalies

Threshold = The 90th quantile of the maximum distances computed by PCA from temporal variation aspect

doi:10.6342/NTU202001579

Table 7.

Nine Categories of Anomalies Detected by PCA Method for Each Storm Type

Code	B1*	B2*	B3	B4	B5*	B6	B7	B8	B9*	Sum
Frontal Rain	50	0	3	2	10	7	2	1	1	76
Meiyu	56	1	9	3	18	5	0	4	5	101
Convective Storms	87	1	4	0	4	0	1	0	4	101
Typhoons	20	1	0	0	0	0	0	0	1	22
Sum	213	3	16	5	32	12	3	5	11	300

Note.
* indicates that this category of anomalies needs further verification;
OBS = Observations;
B1 = OBS at some hours were higher than that of nearby stations;
B2 = Observed trace; nearby stations, rainfalls;
B3 = Did not observe OBS due to malfunctions;
B4 = Did not observe OBS due to delays;
B5 = Observed rainfalls; nearby stations, trace;
B6 = Observed rainfalls; nearby stations did not due to malfunctions;
B7 = Observed rainfalls; nearby stations did not due to delays;
B8 = Delayed return of accumulated rainfall records;
B9 = Rainfall trend was different from that of nearby stations

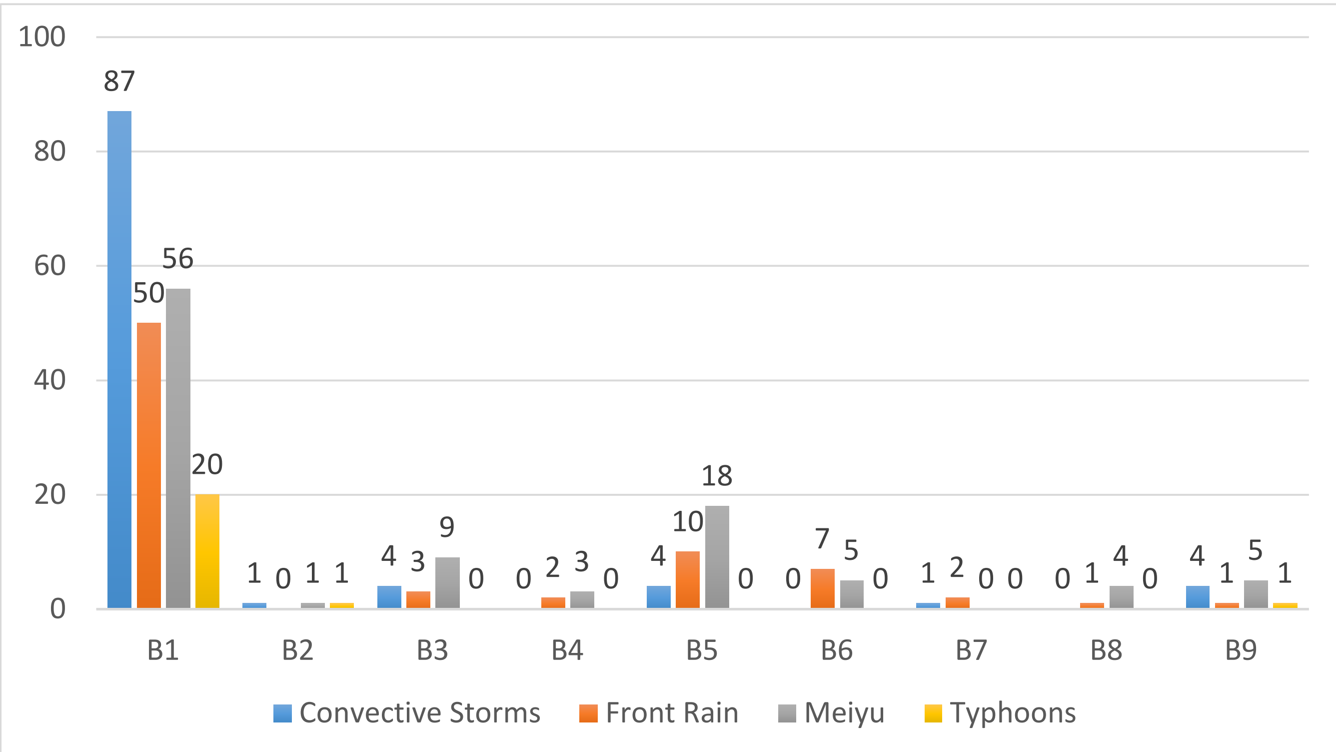


Figure 9. The Bar Chart of Nine Categories of Anomalies of Each Storm Type

doi:10.6342/NTU202001579

The anomaly that occurs at the Donghe Station on March 27, 2020, belongs to category B1. For B2 to B9, we selected one anomaly detected for each category, and it is present as follows:

Category B2

From July 16 to July 20, 2005, Typhoon Haitang struck Taiwan. Xingaokou Station observed very little rainfall on July 18, 2005 (trace), while other three nearby stations did observe relatively high rainfall.

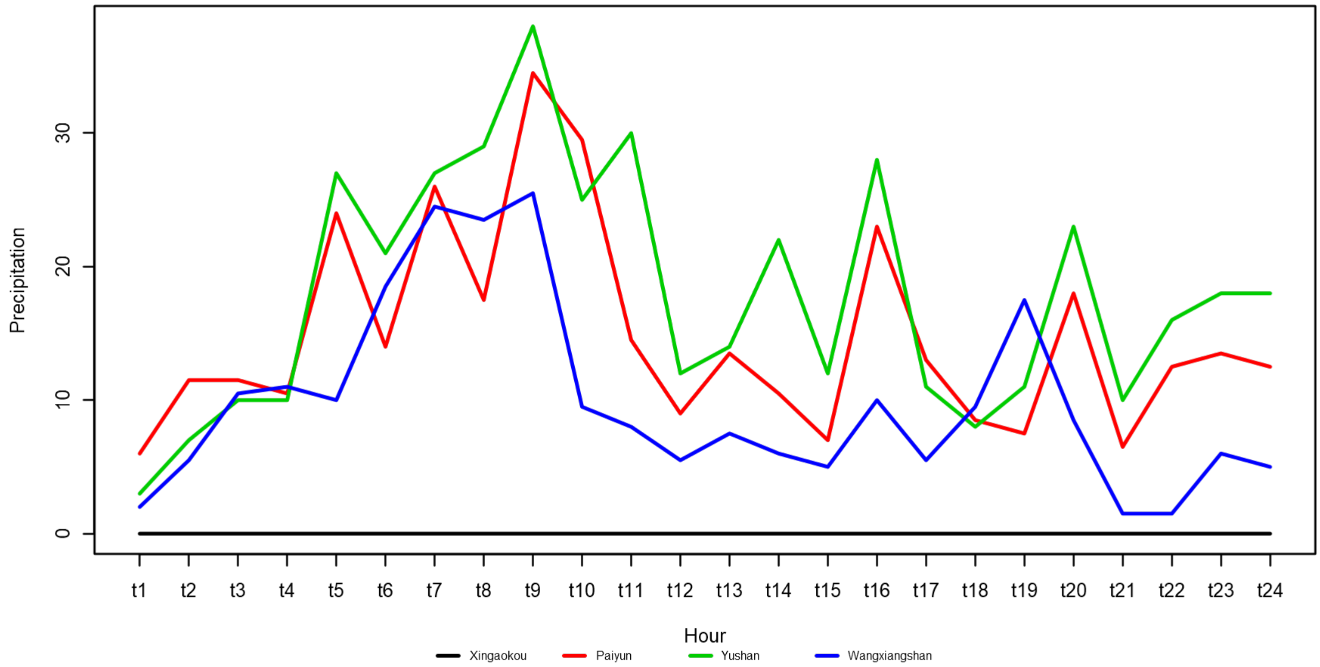


Figure 10. Rainfalls Observed by Xingaokou Station and Nearby Stations on Jul 18, 2005

Xingaokou Station is detected to have anomalies. The neighboring stations are Paiyun, Yushan, and Wangxiangshan, from near to far.

doi:10.6342/NTU202001579

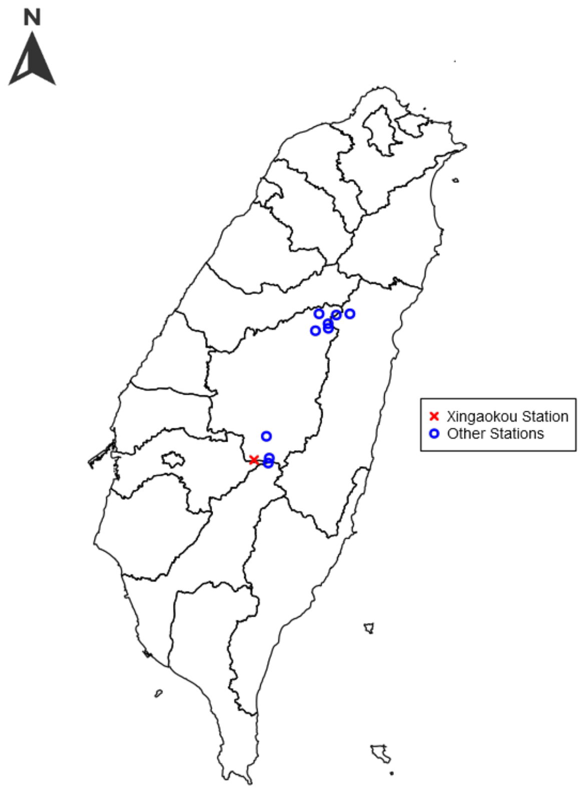


Figure 11. Anomaly Detected on Jul 18, 2005, in Cluster 2 of Typhoon

The red cross represents where Xingaokou Station is located, while there exist anomalies in the rainfalls that Xingaokou Station observed. The blue circles are other rain gauge stations, observing no anomalies, in Cluster 2.

doi:10.6342/NTU202001579

Category B3

Siyuan Station did not observe any rainfall on June 2, 2017, because of malfunctions; whereas other nearby five stations in the same cluster observed rainfalls.

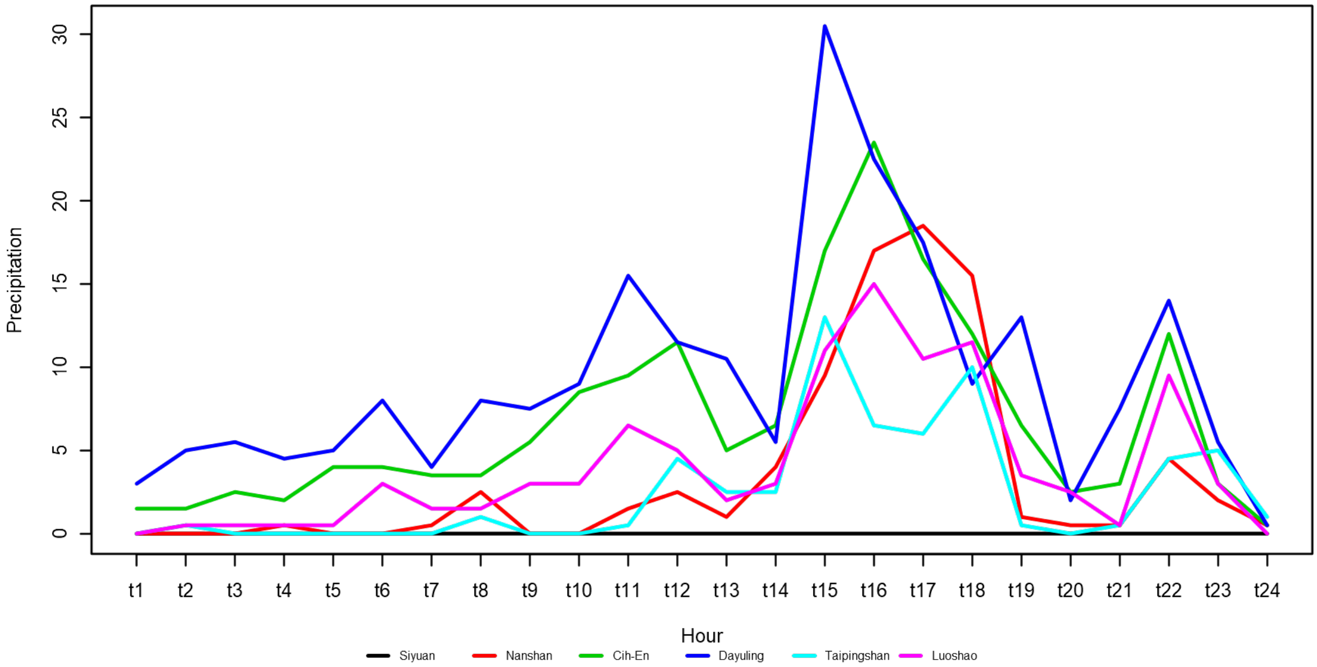


Figure 12. Rainfalls Observed by Siyuan Station and Nearby Stations on Jun 2, 2017

Siyuan Station is detected to have anomalies. The neighboring stations are Nanchan, Cih-En, Dayuling, Taipingshan, and Luoshao, from near to far.

doi:10.6342/NTU202001579

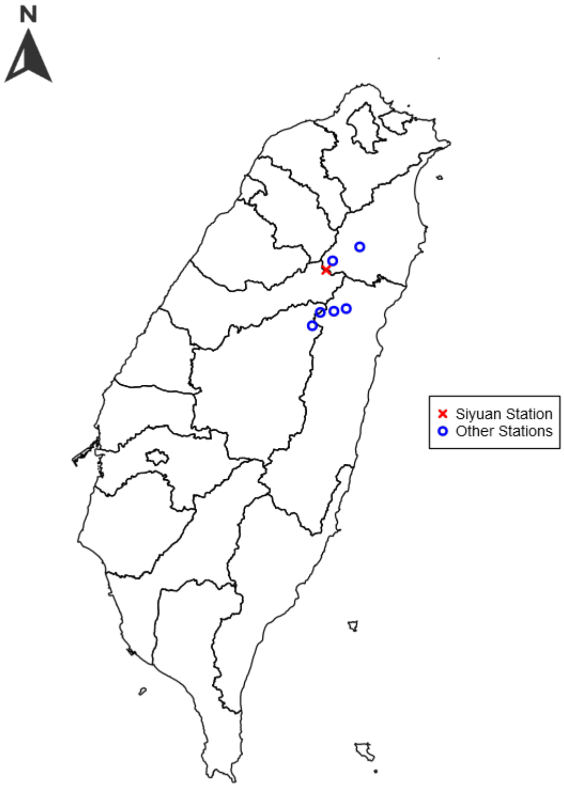


Figure 13. Anomaly Detected on Jun 2, 2017, in Cluster 1 of Meiyu

The red cross represents where Siyuan Station is located, while there exist anomalies in the rainfalls that Siyuan Station observed. The blue circles are other rain gauge stations, observing no anomalies, in Cluster 1.

doi:10.6342/NTU202001579

Category B4

Daping Station did not observe any rainfall on November 27, 1998, because the rain gauge delayed in returning the observations in time, while other nearby two stations did.

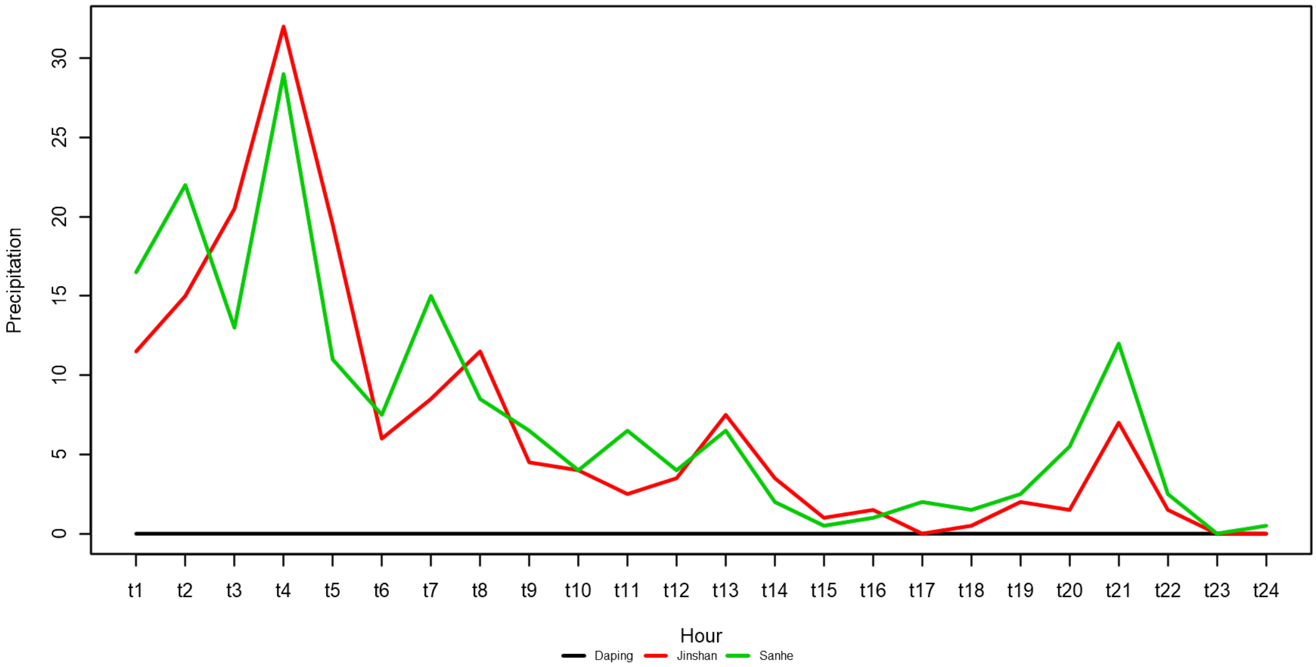


Figure 14. Rainfalls Observed by Daping Station and Nearby Stations on Nov 27, 1998

Daping Station is detected to have anomalies. The neighboring stations are Jinshan and Sanhe, from near to far.

doi:10.6342/NTU202001579

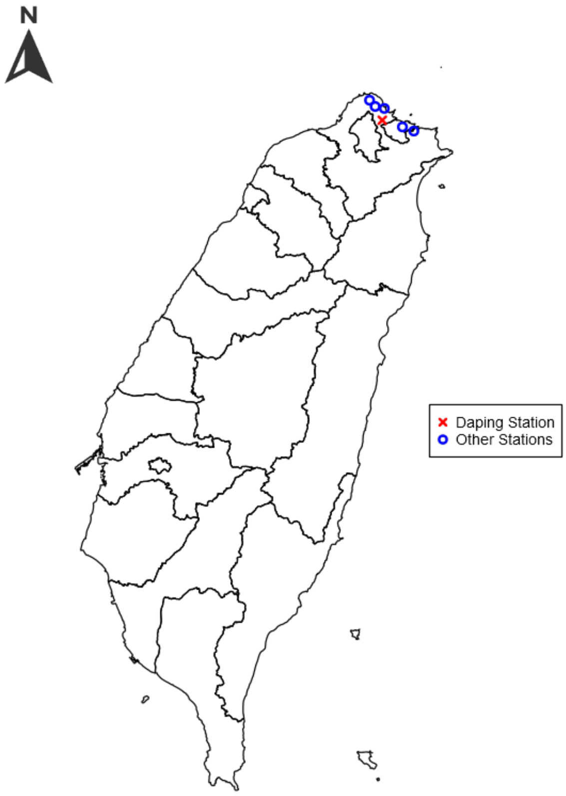


Figure 15. Anomaly Detected on Nov 27, 1998, in Cluster 8 of Frontal Rain

The red cross represents where Daping Station is located, while there exist anomalies in the rainfalls that Daping Station observed. The blue circles are other rain gauge stations, observing no anomalies, in Cluster 8.

doi:10.6342/NTU202001579

Category B5

Only Station Tonemen did observe rainfall on June 17, 1998, while other stations in the same cluster (Cluster 3 of Meiyu) did not.

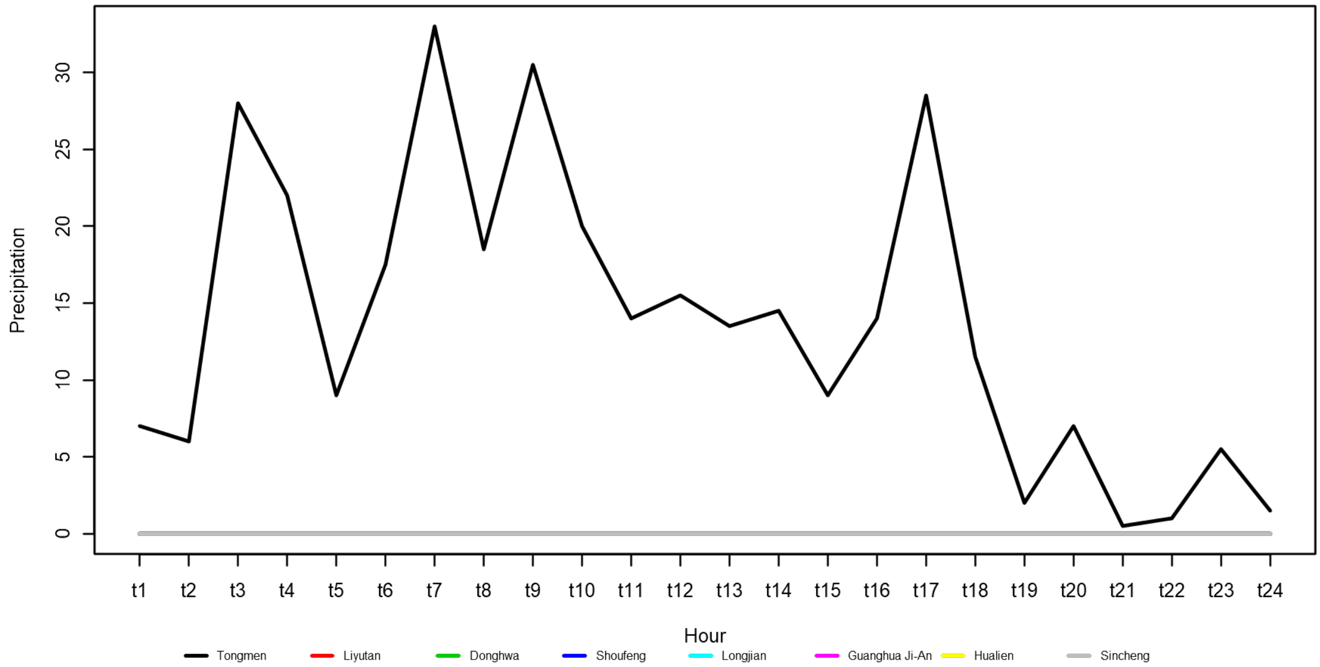


Figure 16. Rainfalls Observed by Tongmen Station and Nearby Stations on Jun 17, 1998

Tongmen Station is detected to have anomalies. The neighboring stations are Liyutan, Donghwa, Shoufeng, Longjian, Guanghua Ji-An, and Hualien, from near to far.

doi:10.6342/NTU202001579

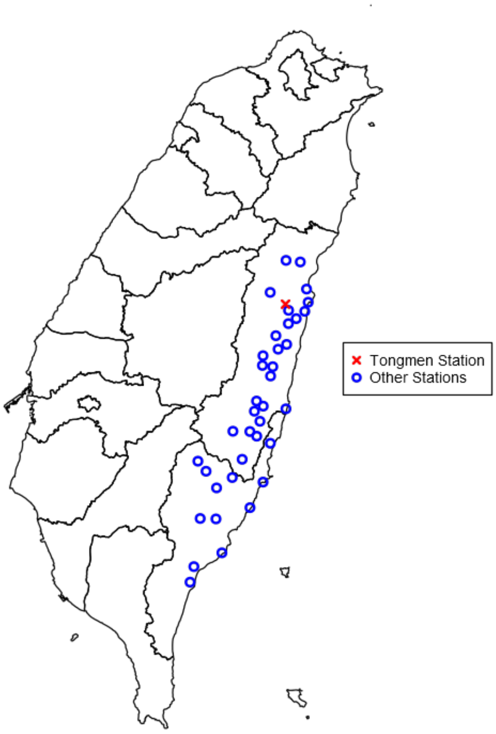


Figure 17. Anomaly Detected on Jun 17, 1998, in Cluster 3 of Meiyu

The red cross represents where Tongmen Station is located, while there exist anomalies in the rainfalls that Tongmen Station observed. The blue circles are other rain gauge stations, observing no anomalies, in Cluster 3.

doi:10.6342/NTU202001579

Category B6

Station Taipingshan did observe rainfall on November 7, 2017, while other nearby stations merely didn’t.

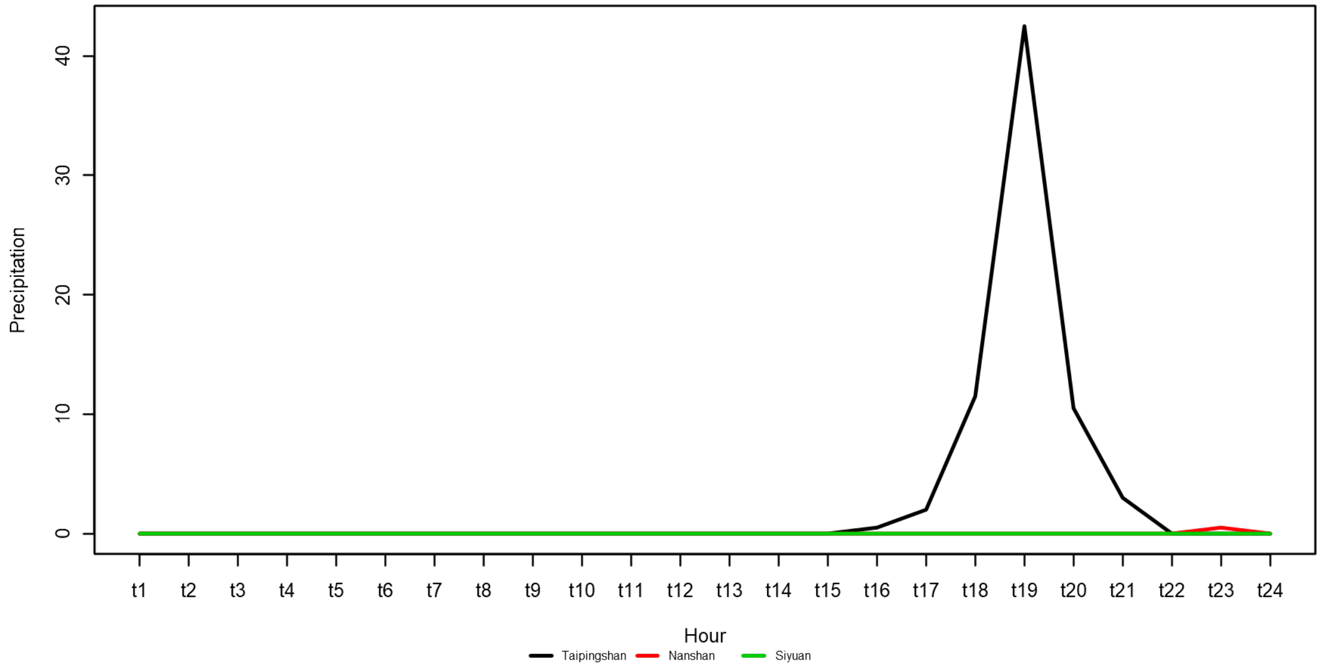


Figure 18. Rainfalls Observed by Taipingshan Station and Nearby Stations on Nov 7, 2017.

Station Taipingshan is detected to have anomalies. The neighboring stations are Nanshan and Siyuan, from near to far.

doi:10.6342/NTU202001579

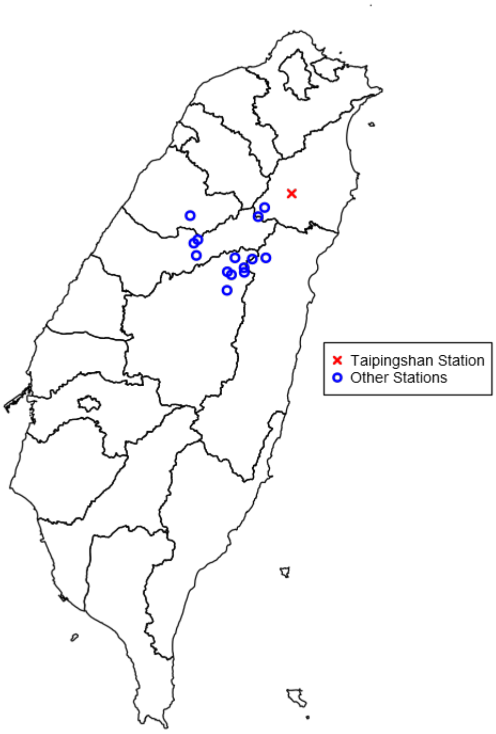


Figure 19. Anomaly Detected on Nov 7, 2017, in Cluster 7 of Frontal Rain

The red cross represents where Taipingshan Station is located, while there exist anomalies in the rainfalls that Taipingshan Station observed. The blue circles are other rain gauge stations, observing no anomalies, in Cluster 7.

doi:10.6342/NTU202001579

Category B7

Station Taitung did observe any rainfall on August 27, 2001, while the other nearby stations didn’t because they delayed in returning their observations.

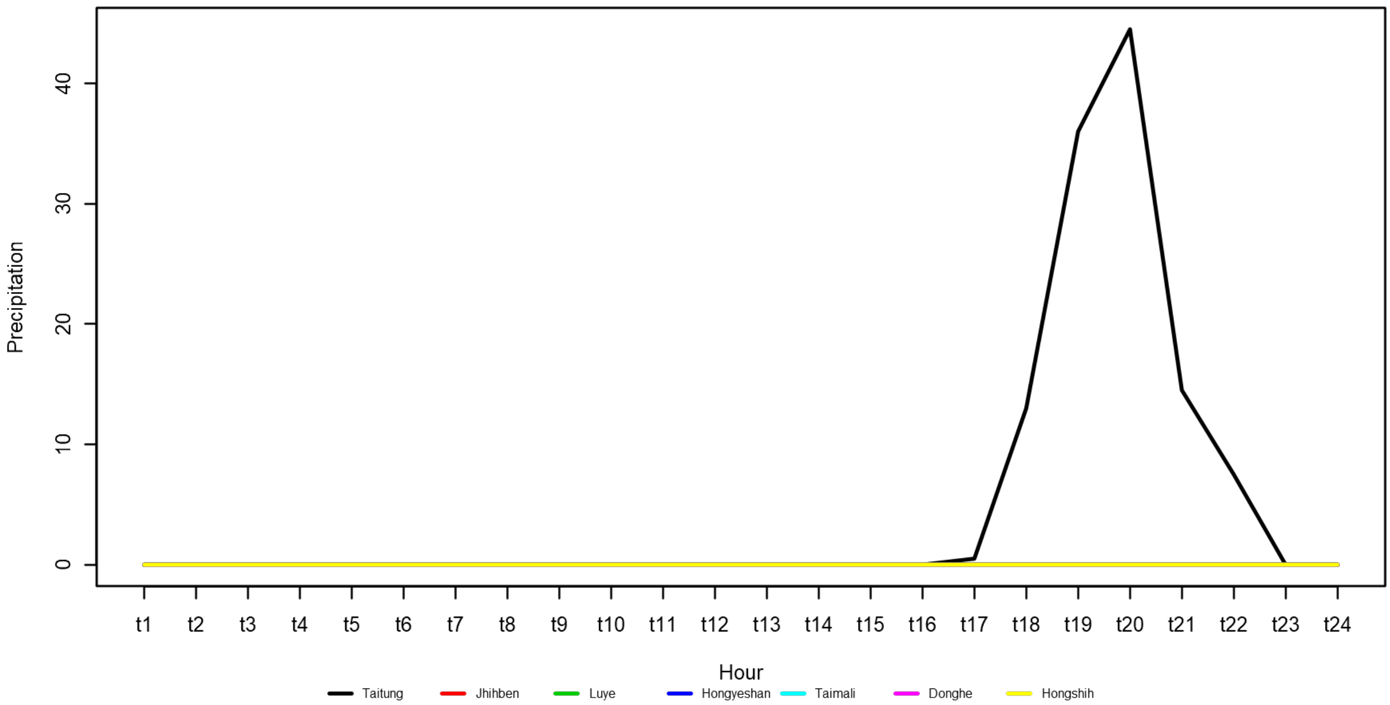


Figure 20. Rainfalls Observed by Taitung Station and Nearby Stations on Aug 27, 2001

Taitung Station is detected to have anomalies. The neighboring stations are Jiben, Luye, Hongyeshan, Taimali, Donghe, and Hongshih, from near to far.

doi:10.6342/NTU202001579

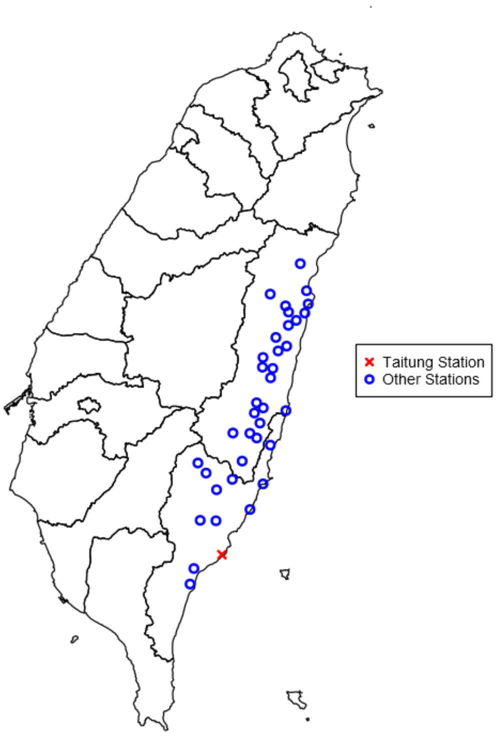


Figure 21. Anomaly Detected on Aug 27, 2001, in Cluster 2 of Convective Storms

The red cross represents where Taitung Station is located, while there exist anomalies in the rainfalls that Taitung Station observed. The blue circles are other rain gauge stations, observing no anomalies, in Cluster 2.

doi:10.6342/NTU202001579

Category B8

Shigang Station delayed returning its observations for several hours on June 25, 1998.

Until 7 pm, it returned 56 mm rainfall and 8 pm it returned 91 mm rainfall.

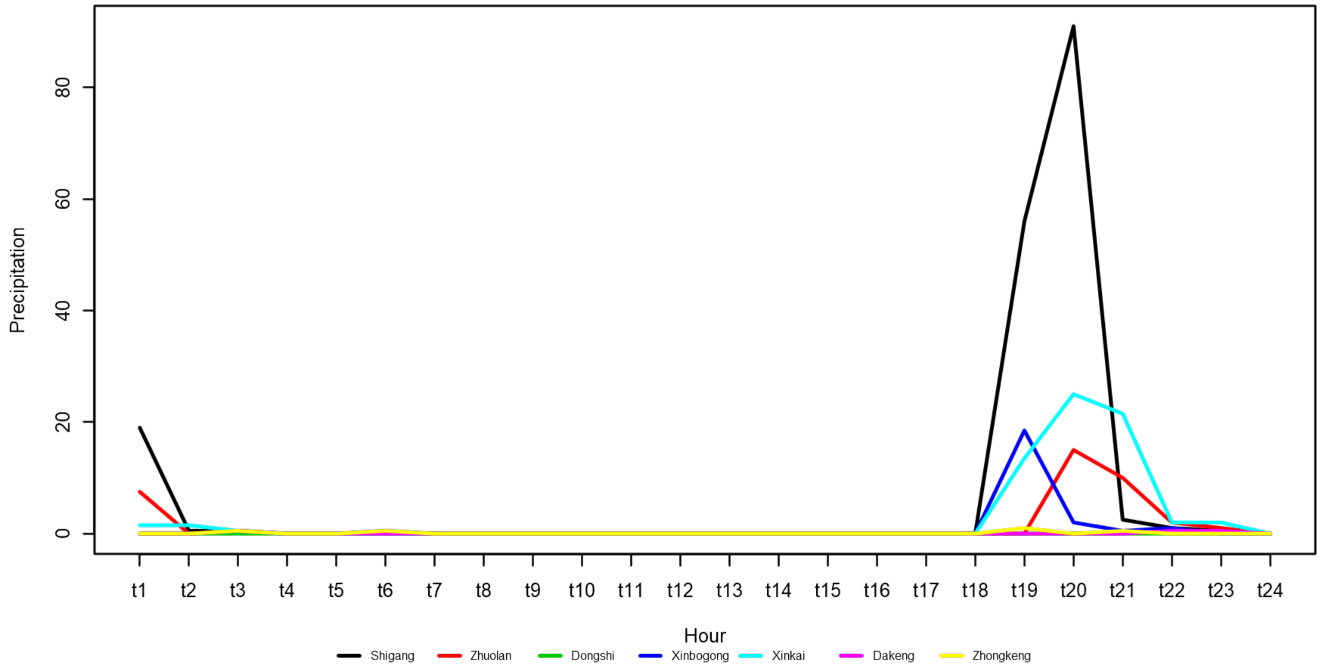


Figure 22. Rainfalls Observed by Shigang Station and Nearby Stations on Jun 25, 1998

Shigang Station is detected to have anomalies. The neighboring stations are Zhuolan, Dongshi, Xinbogong, Xinkai, Dakeng, and Zhongkeng, from near to far.

doi:10.6342/NTU202001579

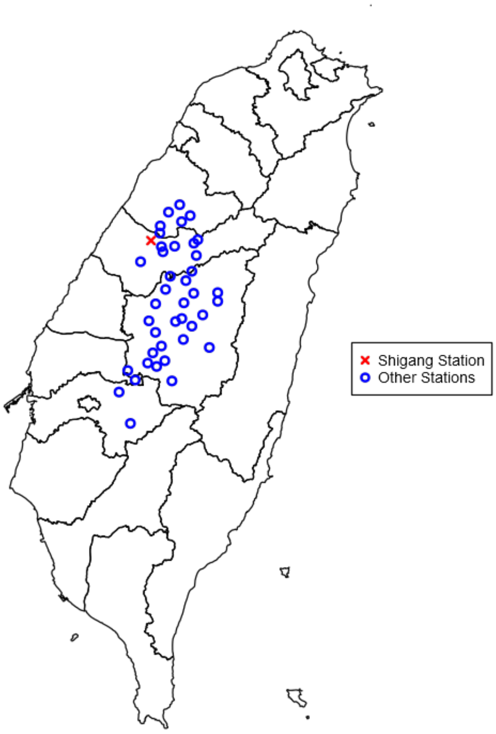


Figure 23. Anomaly Detected on Jun 25, 1998, in Cluster 8 of Meiyu

The red cross represents where Shigang Station is located, while there exist anomalies in the rainfalls that Shigang Station observed. The blue circles are other rain gauge stations, observing no anomalies, in Cluster 8.

doi:10.6342/NTU202001579

Category B9

From Figure 24, the rainfall pattern of Guoxing Station was considerably different from other nearby stations on Jul 10, 1999. Seven nearby stations observed rainfalls from 7 a.m. to 12 p.m., whereas Guoxing Station did not until 6 p.m.

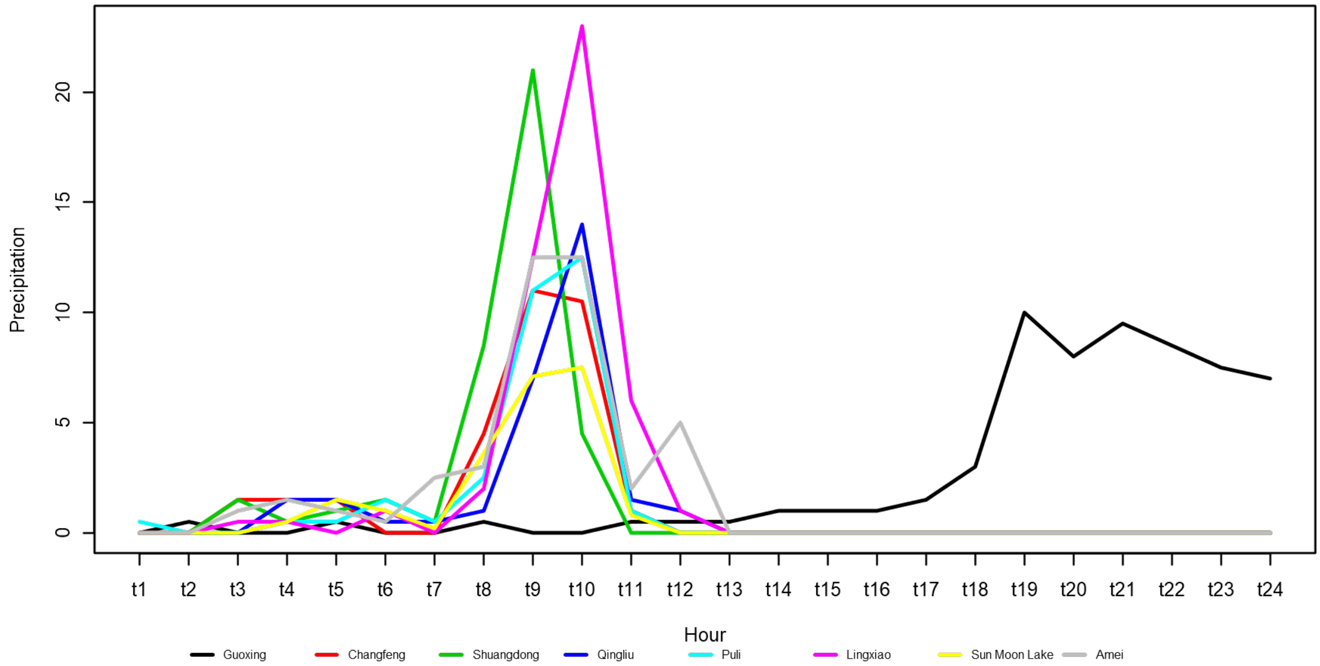


Figure 24. Rainfalls Observed by Guoxing Station and Nearby Stations on Jul 10, 1999

Guoxing Station is detected to have anomalies. The neighboring stations are Changfeng, Shuangdong, Qingliu, Puli, Lingxiao, Sun Moon Lake, and Amei, from near to far.

doi:10.6342/NTU202001579

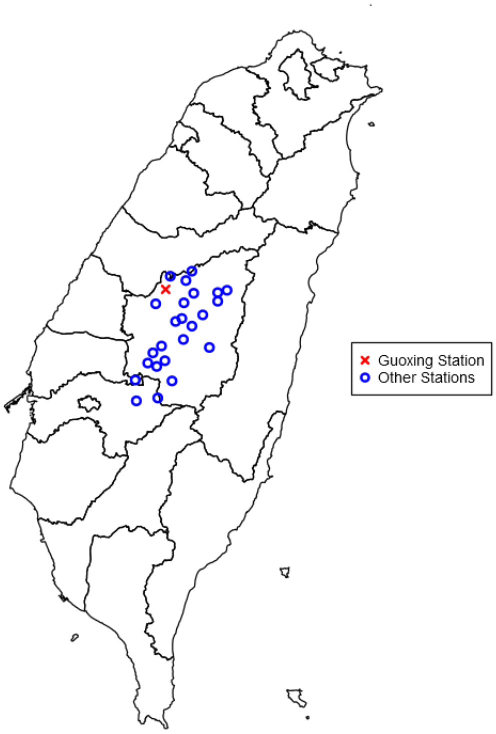


Figure 25. Anomaly Detected on Jul 10, 1999, in Cluster 7 of Convective Storms

The red cross represents where Guoxing Station is located, while there exist anomalies in the rainfalls that Guoxing Station observed. The blue circles are other rain gauge stations, observing no anomalies, in Cluster 7.

doi:10.6342/NTU202001579

4-3-3. The Automated Anomaly Detection System

In this section, we introduce our automated anomaly detection application system. We established the system using Shiny, a package developed by R Studio for users to create interactive web pages with R language. The URL of our online system is https://roam041.shinyapps.io/outlier_detection_v1/.

Two tabs are present in the navigation: Clustering and PCA. The Clustering tab provides users with the clustering results of each storm type of 297 rain gauge stations. On the left-hand side, users need to select the specific storm type (frontal rain, Meiyu, convective storms, or typhoons) and the cluster from the two dropdown select options. Once the options are selected, the image box displays the clustering result of the chosen storm type, and the map box displays an interactive map with detailed information of rain gauge stations of a determined cluster.

The PCA tab presents the results using the PCA method. In the Options box, users select the preferred storm type, cluster, and the day (of all available days). Then, the Map box on the right-hand side reveals the anomaly detection result of the chosen day. The red pins with an exclamation mark indicates locations of anomalies, whereas the blue pins

doi:10.6342/NTU202001579

with a check sign indicate locations where no anomalies exist. The Precipitation box shows the rainfall of the chosen date (24 h). Users can even draw the interactive rainfall time series plot if they select stations in the Chosen Stations input box. Finally, the tab presents the PCA results from both spatial and temporal variation aspects: the variable correlation plot, the individuals' plot (in the new transformed coordinate system), and the loadings of all variables.

doi:10.6342/NTU202001579

Conclusion

We established an automated anomaly detection system for hourly precipitation data. Anomalies can occur at a station because of two main abnormal situations. First, the station fails to return the rainfall observations in time because of malfunctions or delays, which results in an exceptionally high rainfall since because of data accumulation for several hours. We developed the cutoff point method to detect this type of anomalies.

Second, the rainfall observed by the station is extraordinarily higher or lower than that of the neighboring stations. We adopted the K-means cluster analysis to group the 297 stations based on geographical locations and rainfall characteristics as per the four primary storm types in Taiwan. Then, we used PCA to compute d, the Euclidean distance of the projected data point from the origin for each observation. When hours were taken as variables for PCA, d represented the temporal variation of the rainfall at each station in a specified cluster. By contrast, d represented for the spatial variation of the rainfall patterns in different hours when each variable contained the hourly rainfalls at a station within a day. When the value of d exceeded the threshold set, our system automatically indicates possible anomalies. The anomalies identified with the PCA method have nine categories. Some of them may not be anomalies, which still require

doi:10.6342/NTU202001579

additional verification. Nevertheless, our system can effectively and efficiently screen out the potential anomalies to achieve the QC of hourly rainfall data.

doi:10.6342/NTU202001579

References

Bock, H. (1985). On some significance tests in cluster analysis. *Journal of Classification*, *2*(1), 77-108. doi: 10.1007/bf01908065

Bock, H. (2008). Origins and extensions of the k-means algorithm in cluster analysis. *Electronic Journal for History of Probability and Statistics*, *4*(2), 18.

Boyle, J. S., and G. T. J. Chen, (1987). Synoptic aspects of the wintertime East Asian monsoon. *Monsoon Meteorology*, C. P. Chang and T. N. Krishnamurti, Eds., Oxford University Press, 125–160.

Branisavljević, N., Prodanović, D., Arsić, M., Simić, Z., & Borota, J. (2009). Hydro-Meteorological Data Quality Assurance and Improvement. *Journal of The Serbian Society for Computational Mechanics*, *3*(1), 228-249.

Chen, C., & Chen, Y. (2003). The Rainfall Characteristics of Taiwan. *Monthly Weather Review*, *131*(7), 1323-1341. doi: 10.1175/1520-0493(2003)131<1323: trcot>2.0.co;2

Chen, C., & Huang, J. (1999). A Numerical Study of Precipitation Characteristics over Taiwan Island during the Winter Season. *Meteorology and Atmospheric Physics*, *70*(3-4), 167-183. doi: 10.1007/s007030050032

Cox, D. (1957). Note on Grouping. *Journal of the American Statistical Association*, *52*(280), 543-547. doi: 10.1080/01621459.1957.10501411

Fisher, W. (1958). On Grouping for Maximum Homogeneity. *Journal of the American Statistical Association*, *53*(284), 789-798. doi: 10.1080/01621459.1958.10501479

Hartigan, J. (1975). *Clustering algorithms*. New York: John Wiley & Sons.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, *24*(6&7), 417-441 & 498-520. doi: 10.1037/h0071325

Hudson, H., Mcmillan, D., & Pearson, C. (1999). Quality assurance in hydrological measurement. *Hydrological Sciences Journal*, *44*(5), 825-834. doi: 10.1080/02626669909492276

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2015). *An Introduction to Statistical Learning* (6th ed., p. 426). New York: Springer Science + Business.

doi:10.6342/NTU202001579

Jolliffe, I. (2002). *Principal Component Analysis. 2nd ed* (2nd ed.). New York, NY: Springer-Verlag.

Kassambara, A. (2020). factoextra package | R Documentation. Retrieved 29 May 2020, from https://www.rdocumentation.org/packages/factoextra

Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *2*(11), 559-572. doi: 10.1080/14786440109462720

Pollard, D. (1982). A Central Limit Theorem for K-Means Clustering. *The Annals of Probability*, *10*(4), 919-926. doi: 10.1214/aop/1176993713

Rousseeuw, P., & Hubert, M. (2018). Anomaly detection by robust statistics. *Wires Data Mining and Knowledge Discovery*, *8*(2). doi: 10.1002/widm.1236

Toe, M., Kanzaki, M., Lien, T., & Cheng, K. (2017). Spatial and temporal rainfall patterns in Central Dry Zone, Myanmar - A hydrological cross-scale analysis. *Terrestrial, Atmospheric and Oceanic Sciences*, *28*(3), 425-436. doi: 10.3319/tao.2016.02.15.01(Hy)

Wang, S.-T., and H. Cheng, (1982). Natural seasons as shown by the variation of the circulation in Asia (in Chinese). *Atmos. Sci.*, 9, 125–146.

You, J., Hubbard, K., Nadarajah, S., & Kunkel, K. (2007). Performance of Quality Assurance Procedures on Daily Precipitation. *Journal of Atmospheric and Oceanic Technology*, *24*(5), 821-834. doi: 10.1175/jtech2002.1

doi:10.6342/NTU202001579

Appendix

A. Table of 297 Stations

Item	ID	Station	站名	Altitude	Longitude	Latitude	C.	F.	M.	T.
1	466900	Tamsui	淡水	19	121.449	25.1649	8	3	6	4
2	466910	Anbu	鞍部	825.8	121.530	25.1826	8	3	6	4
3	466920	Taipei	臺北	6.3	121.515	25.0377	8	3	6	4
4	466930	Zhuzihu	竹子湖	607.1	121.545	25.1621	8	3	6	4
5	466940	Keelung	基隆	26.7	121.741	25.1333	8	8	6	4
6	466990	Hualien	花蓮	16	121.613	23.9751	2	4	3	1
7	467060	Su-Ao	蘇澳	24.9	121.857	24.5967	4	3	6	5
8	467080	Yilan	宜蘭	7.2	121.757	24.7640	4	3	6	5
9	467410	Tainan	臺南	40.8	120.205	22.9932	6	5	4	6
10	467420	Yongkang	永康	8.1	120.237	23.0384	6	5	4	7
11	467440	Kaohsiung	高雄	2.3	120.316	22.5660	6	5	4	7
12	467480	Chiayi	嘉義	26.9	120.433	23.4959	1	1	2	6
13	467490	Taichung	臺中	84	120.684	24.1457	3	6	2	6
14	467530	Alishan	阿里山	2413.4	120.813	23.5082	7	2	5	8
15	467540	Dawu	大武	8.1	120.904	22.3557	6	5	4	7
16	467550	Yushan	玉山	3844.8	120.960	23.4876	10	2	5	2
17	467571	Hsinchu	新竹	26.9	121.014	24.8279	8	6	6	4
18	467590	Hengchun	恆春	22.1	120.746	22.0039	6	5	4	7
19	467610	Chenggong	成功	33.5	121.373	23.0975	2	4	3	1
20	467650	Sun Moon Lake	日月潭	1017.5	120.908	23.8813	7	6	8	3
21	467660	Taitung	臺東	9	121.155	22.7522	2	4	3	1
22	467770	Wuqi	梧棲	31.7	120.523	24.2560	1	1	2	6
23	C0A520	Shanjia	山佳	48	121.402	24.9749	8	3	6	4
24	C0A530	Pinglin	坪林	300	121.709	24.9382	8	3	6	5
25	C0A540	Sihdu	四堵	401	121.746	24.8928	4	3	6	5
26	C0A550	Taiping	泰平	422	121.824	24.9712	4	3	6	5
27	C0A570	Tonghou	桶後	360	121.598	24.8482	4	3	6	5
28	C0A640	Shihding	石碇	241	121.663	24.9939	8	3	6	5
29	C0A650	Huoshaoliao	火燒寮	287	121.743	25.0027	8	3	6	5
30	C0A660	Rueifang	瑞芳	97	121.801	25.1132	8	8	6	4

Note. C., F., M., T.=The clusters of convective storms, frontal rain, meiyu, typhoons.

doi:10.6342/NTU202001579

A. Table of 297 Stations (Cont'd)

Item	ID	Station	站名	Altitude	Longitude	Latitude	C.	F.	M.	T.
31	C0A860	Daping	大坪	362	121.633	25.1659	8	8	6	4
32	C0A870	Wujhihshan	五指山	685	121.609	25.1322	8	3	6	4
33	C0A880	Fulong	福隆	6	121.942	25.0178	4	3	6	5
34	C0A890	Shuangsi	雙溪	40	121.864	25.036	8	3	6	5
35	C0A920	Fugueijiao	富貴角	196	121.565	25.2638	8	8	6	4
36	C0A931	Sanhe	三和	216	121.595	25.2332	8	8	6	4
37	C0A940	Jinshan	金山	49	121.644	25.2236	8	8	6	4
38	C0A970	Sandiaojiao	三貂角	96	122.002	25.0076	4	3	6	5
39	C0A980	Shezih	社子	11	121.47	25.1095	8	3	6	4
40	C0A9A0	Dazhi	大直	24	121.543	25.078	8	3	6	4
41	C0A9C0	Tianmu	天母	35	121.537	25.1175	8	3	6	4
42	C0A9E0	Shihlin	士林	26	121.503	25.0903	8	3	6	4
43	C0A9F0	Neihu	內湖	35	121.576	25.0794	8	3	6	4
44	C0AC80	Wenshan	文山	40	121.576	25.0024	8	3	6	4
45	C0ACA0	Xinzhuang	新莊	25	121.447	25.0515	8	3	6	4
46	C0AG90	Zhonghe	中和	25	121.49	24.9926	8	3	6	4
47	C0AH10	Yonghe	永和	30	121.508	25.0113	8	3	6	4
48	C0AH40	Pingdeng	平等	426	121.577	25.1291	8	3	6	4
49	C0AH50	Linkou	林口	275	121.381	25.0722	8	3	6	4
50	C0AI10	Cyuchih	屈尺	76	121.545	24.9218	8	3	6	5
51	C0AI40	Shipai	石牌	35	121.513	25.1156	8	3	6	4
52	C0C460	Fuxing	復興	482	121.352	24.8202	8	3	6	4
53	C0C480	Taoyuan	桃園	105	121.323	24.9924	8	3	6	4
54	C0C490	Bade	八德	157	121.283	24.9287	8	3	6	4
55	C0C540	Dayuan	大園	46	121.226	25.0478	8	3	6	4
56	C0C660	Yangmei	楊梅	176	121.143	24.9124	8	6	6	4
57	C0C700	Zhongli (NCU)	中壢	151	121.256	24.9777	8	3	6	4
58	C0D360	Meihua	梅花	523	121.209	24.6783	3	6	6	4
59	C0D390	Guanxi	關西	146	121.174	24.7982	3	6	6	4
60	C0D430	Emei	峨眉	87	121.017	24.6905	3	6	6	4

Note. C., F., M., T.=The clusters of convective storms, frontal rain, meiyu, typhoons.

doi:10.6342/NTU202001579

A. Table of 297 Stations (Cont'd)

Item	ID	Station	站名	Altitude	Longitude	Latitude	C.	F.	M.	T.
61	C0D480	Datiekeng	打鐵坑	223	121.151	24.8497	8	6	6	4
62	C0D650	Hukou	湖口	97	121.044	24.9048	8	6	6	4
63	C0E420	Jhunan	竹南	19	120.889	24.7090	1	6	6	4
64	C0E430	Nanzhuang	南庄	258	121.000	24.6018	3	6	6	3
65	C0E520	Dahu	大湖	320	120.871	24.4149	3	6	8	3
66	C0E540	Houlong	後龍	32	120.784	24.6065	1	6	2	6
67	C0E550	Mingde	明德	84	120.885	24.5835	3	6	6	4
68	C0E590	Tongxiao	通霄	40	120.705	24.4719	1	6	2	6
69	C0E610	Madu-An	馬都安	850	120.930	24.4513	3	6	8	3
70	C0E790	Zhuolan	卓蘭	344	120.825	24.3128	3	6	8	3
71	C0E820	Shitan	獅潭	220	120.920	24.5391	3	6	6	3
72	C0E830	Yuanli	苑裡	37	120.653	24.4397	1	6	2	6
73	C0E850	Dahe	大河	104	120.949	24.6181	3	6	6	4
74	C0F000	Dadu	大肚	273	120.572	24.1530	1	1	2	6
75	C0F0B0	Shigang	石岡	311	120.778	24.2761	3	6	8	3
76	C0F0C0	Zhongkeng	中坑	983	120.903	24.2485	3	6	8	3
77	C0F850	Dongshi	東勢	379	120.833	24.2464	3	6	8	3
78	C0F930	Dajia	大甲	100	120.640	24.3477	1	1	2	6
79	C0F970	Dakeng	大坑	145	120.722	24.1731	3	6	8	6
80	C0F9X0	Daya	大雅（中科園區）	166	120.625	24.2153	1	1	2	6
81	C0G640	Lukang	鹿港	17	120.431	24.0753	1	1	2	6
82	C0G650	Yuanlin	員林	34	120.586	23.9465	1	1	2	6
83	C0G660	Xihu	溪湖	27	120.479	23.9484	1	1	2	6
84	C0G720	Xizhou	溪州	50	120.499	23.8516	1	1	2	6
85	C0G730	Erlin	二林	27	120.376	23.9033	1	1	2	6
86	C0G740	Dacheng	大城	24	120.321	23.8525	1	1	2	6
87	C0H890	Puli	埔里	440	120.952	23.9722	7	6	8	3
88	C0H950	Zhongliao	中寮	192	120.766	23.8844	5	1	8	6
89	C0H960	Caotun	草屯	120	120.681	23.9737	3	1	2	6
90	C0H990	Kunyang	昆陽	3076	121.274	24.1213	9	7	1	2

Note. C., F., M., T.=The clusters of convective storms, frontal rain, meiyu, typhoons.

doi:10.6342/NTU202001579

A. Table of 297 Stations (Cont'd)

Item	ID	Station	站名	Altitude	Longitude	Latitude	C.	F.	M.	T.
91	C0H9C0	Hehuan Mountain	合歡山	3402	121.273	24.1434	9	7	5	2
92	C0I010	Lushan	廬山	1562	121.182	24.0333	7	7	5	3
93	C0I080	Xinyi	信義	536	120.851	23.6897	7	2	8	3
94	C0I090	Fonghuang	鳳凰	910	120.787	23.7281	7	2	8	3
95	C0I110	Zhushan	竹山	161	120.688	23.7612	5	1	2	6
96	C0I370	Yuchi	魚池	671	120.941	23.8957	7	6	8	3
97	C0I380	Jiji	集集	258	120.802	23.8282	5	1	8	6
98	C0I390	Ren'Ai	仁愛	1184	121.132	24.0221	7	6	8	3
99	C0I420	Guoxing	國姓	305	120.855	24.0378	7	6	8	3
100	C0K240	Caoling	草嶺	1132	120.694	23.5956	7	2	8	8
101	C0K250	Lunbei	崙背	12	120.319	23.7556	1	1	2	6
102	C0K280	Sihu	四湖	23	120.227	23.6304	1	1	2	6
103	C0K291	Yiwu	宜梧	8	120.169	23.5363	1	1	2	6
104	C0K330	Huwei	虎尾	38	120.442	23.7192	1	1	2	6
105	C0K390	Tuku	土庫	31	120.396	23.6790	1	1	2	6
106	C0K400	Douliu	斗六	65	120.541	23.7206	1	1	2	6
107	C0K410	Beigang	北港	20	120.293	23.5740	1	1	2	6
108	C0K420	Xiluo	西螺	42	120.467	23.8004	1	1	2	6
109	C0K430	Baozhong	褒忠	30	120.304	23.6909	1	1	2	6
110	C0K460	Dounan	斗南	60	120.478	23.6787	1	1	2	6
111	C0K490	Gukeng	古坑	91	120.560	23.6543	5	1	2	6
112	C0M410	Matoushan	馬頭山	245	120.582	23.3244	5	1	7	7
113	C0M520	Donghouliao	東後寮	15	120.248	23.3699	1	1	2	6
114	C0M530	Fenqihu	奮起湖	1385	120.699	23.4939	7	2	7	8
115	C0M640	Zhongpu	中埔	155	120.523	23.4254	5	1	2	6
116	C0M650	Puzi	朴子	20	120.239	23.4346	1	1	2	6
117	C0M660	Xikou	溪口	40	120.404	23.6041	1	1	2	6
118	C0M680	Taibao	太保	37	120.332	23.4551	1	1	2	6
119	C0M690	Shuishang	水上	33	120.389	23.4197	1	1	2	6
120	C0M700	Zhuqi	竹崎	150	120.556	23.5262	5	1	2	6

Note. C., F., M., T.=The clusters of convective storms, frontal rain, meiyu, typhoons.

doi:10.6342/NTU202001579

A. Table of 297 Stations (Cont'd)

Item	ID	Station	站名	Altitude	Longitude	Latitude	C.	F.	M.	T.
121	C0M710	Dongshi	東石	15	120.154	23.4589	1	1	2	6
122	C0M770	Meishan Chiayi County	嘉義梅山	164	120.556	23.5854	5	1	2	6
123	C0M830	Shanmei	山美	540	120.668	23.3838	5	2	8	8
124	C0O810	Cengwen	曾文	161	120.497	23.2197	5	1	4	7
125	C0O830	Beiliao	北寮	127	120.495	23.0796	5	5	4	7
126	C0O840	Wangyegong	王爺宮	134	120.401	23.2221	6	1	4	7
127	C0O860	Danei	大內	38	120.361	23.1189	6	1	4	7
128	C0O900	Shanhua	善化	9	120.297	23.1129	6	1	4	6
129	C0O930	Yujing	玉井	69	120.461	23.1260	6	1	4	7
130	C0O950	Annan	安南	4	120.145	23.0767	6	1	4	6
131	C0O960	Qiding	崎頂	112	120.369	22.9595	6	5	4	7
132	C0O970	Hutoupi	虎頭埤	71	120.348	23.0214	6	5	4	7
133	C0O980	Xinshi	新市	18	120.298	23.0616	6	5	4	7
134	C0O990	Mamiao	媽廟	18	120.294	22.9918	6	5	4	7
135	C0R130	Ali	阿禮	1040	120.744	22.7429	5	5	7	8
136	C0R140	Majia	瑪家	740	120.687	22.6829	5	5	7	8
137	C0R150	Sandimen	三地門	105	120.640	22.7099	5	5	7	7
138	C0R160	Yanpuxinwei	鹽埔新圍	45	120.531	22.7396	5	5	4	7
139	C0R170	Pingdong	屏東	26	120.494	22.6603	5	5	4	7
140	C0R190	Chishan	赤山	32	120.614	22.5923	5	5	7	7
141	C0R220	Chaojhou	潮州	23	120.540	22.5344	6	5	4	7
142	C0R240	Laiyi	來義	87	120.625	22.5273	5	5	7	7
143	C0R260	Chunri	春日	76	120.628	22.3704	6	5	4	7
144	C0R280	Binlang	檳榔	242	120.837	22.0761	6	5	4	7
145	C0R320	Checheng	車城	7	120.716	22.0740	6	5	4	7
146	C0R341	Mudan	牡丹	230	120.793	22.1300	6	5	4	7
147	C0R350	Maobitou	貓鼻頭	35	120.736	21.9218	6	5	4	7
148	C0R420	Mudanchihshan	牡丹池山	504	120.841	22.1678	6	5	4	7
149	C0R570	Linluo	麟洛	37	120.527	22.6508	5	5	7	7
150	C0R580	Nanzhou	南州	10	120.503	22.4859	6	5	4	7

Note. C., F., M., T.=The clusters of convective storms, frontal rain, meiyu, typhoons.

doi:10.6342/NTU202001579

A. Table of 297 Stations (Cont'd)

Item	ID	Station	站名	Altitude	Longitude	Latitude	C.	F.	M.	T.
151	C0R590	Ligang	里港	72	120.495	22.7792	6	5	4	7
152	C0S660	Siama	下馬	794	121.070	23.1504	2	4	3	1
153	C0S680	Hongyeshan	紅葉山	1659	121.039	22.9198	2	4	3	1
154	C0S690	Taimali	太麻里	522	120.985	22.6090	2	5	3	7
155	C0S700	Jhihben	知本	507	121.006	22.6849	2	5	3	7
156	C0S710	Luye	鹿野	382	121.123	22.9177	2	4	3	1
157	C0S740	Chihshang	池上	289	121.210	23.1196	2	4	3	1
158	C0S750	Siangyang	向陽	2280	120.986	23.2484	10	2	5	8
159	C0S760	Hongshih	紅石	1621	121.126	23.0691	2	4	3	1
160	C0S770	Dasishan	大溪山	375	120.943	22.4785	6	5	4	7
161	C0S810	Donghe	東河	65	121.304	22.9727	2	4	3	1
162	C0S830	Changbin	長濱	288	121.412	23.2868	2	4	3	1
163	C0T790	Dayuling	大禹嶺	2830	121.316	24.1861	9	7	1	2
164	C0T820	Tiansiang	天祥	550	121.496	24.1796	9	4	3	1
165	C0T870	Liyutan	鯉魚潭	135	121.509	23.9356	2	4	3	1
166	C0T900	Xilin	西林	160	121.442	23.8119	2	4	3	1
167	C0T960	Guangfu	光復	120	121.425	23.6607	2	4	3	1
168	C0T9M0	Jingpu	靜浦	92	121.495	23.4552	2	4	3	1
169	C0U520	Shuanglianpi	雙連埤	517	121.641	24.7530	4	3	6	5
170	C0U600	Chiaoshi	礁溪	10	121.766	24.8175	4	3	6	5
171	C0U650	Yulan	玉蘭	442	121.587	24.6753	4	3	6	5
172	C0U710	Taipingshan	太平山	1942	121.526	24.5055	9	7	1	3
173	C0U720	Nanshan	南山	1260	121.382	24.4374	9	7	1	3
174	C0U860	Toucheng	頭城	5	121.831	24.8532	4	3	6	5
175	C0U870	Dajiaosi	大礁溪	474	121.675	24.7910	4	3	6	5
176	C0U890	Sansing	三星	116	121.653	24.6681	4	3	6	5
177	C0U900	Neicheng	內城	63	121.688	24.7181	4	3	6	5
178	C0U910	Dongshan	冬山	17	121.794	24.6337	4	3	6	5
179	C0U940	Luodong	羅東	25	121.749	24.6818	4	3	6	5
180	C0V210	Fuxing	復興	734	120.806	23.2224	10	2	7	8

Note. C., F., M., T.=The clusters of convective storms, frontal rain, meiyu, typhoons.

doi:10.6342/NTU202001579

A. Table of 297 Stations (Cont'd)

Item	ID	Station	站名	Altitude	Longitude	Latitude	C.	F.	M.	T.
181	C0V250	Jiasian	甲仙	298	120.592	23.0801	5	5	7	8
182	C0V260	Yuemei	月眉	126	120.54	22.9710	5	5	4	7
183	C0V310	Meinong	美濃	46	120.519	22.8987	5	5	4	7
184	C0V350	Xipu	溪埔	36	120.447	22.7386	6	5	4	7
185	C0V360	Neimen	內門	94	120.467	22.9755	6	5	4	7
186	C0V370	Gutingkeng	古亭坑	74	120.402	22.8932	6	5	4	7
187	C0V400	Agongdian	阿公店	56	120.356	22.8042	6	5	4	7
188	C0V440	Fengshan	鳳山	27	120.356	22.6467	6	5	4	7
189	C0V450	Fengsen	鳳森	51	120.393	22.5451	6	5	4	7
190	C0V660	Gangshan	岡山	25	120.295	22.7971	6	5	4	7
191	C0V730	Daliao	大寮	33	120.396	22.6056	6	5	4	7
192	C0V740	Qishan	旗山	60	120.484	22.8886	5	5	4	7
193	C0V750	Luzhu	路竹	50	120.259	22.8550	6	5	4	7
194	C0V770	Dashe	大社	26	120.347	22.7303	6	5	4	7
195	C0V810	Zuoying	左營	32	120.285	22.6749	6	5	4	7
196	C0X050	Donghe	東河	25	120.386	23.2966	6	1	4	7
197	C0X060	Xiaying	下營	19	120.256	23.2270	6	1	4	6
198	C0X080	Jiali	佳里	2	120.145	23.1730	1	1	2	6
199	C0X200	Zuozhen	左鎮	61	120.409	23.0568	6	5	4	7
200	C0X210	Baihe	白河	38	120.414	23.3475	1	1	4	6
201	C0X230	Yanshui	鹽水	29	120.248	23.2726	1	1	2	6
202	C0X240	Guanziling	關子嶺	394	120.508	23.3312	5	1	7	7
203	C0X250	Xinying	新營	33	120.317	23.3107	1	1	2	6
204	C0X290	Beimen	北門	10	120.126	23.2678	1	1	2	6
205	C0X310	Qigu	七股	9	120.086	23.1473	1	1	2	6
206	C0Z020	Mingli	明里	211	121.262	23.2087	2	4	3	1
207	C0Z050	Jiaxin	佳心	824	121.213	23.3457	2	4	3	1
208	C0Z061	Yuli	玉里	174	121.340	23.3215	2	4	3	1
209	C0Z070	Wuhe	舞鶴	247	121.374	23.4681	2	4	3	1
210	C0Z100	Donghwa	東華	36	121.550	23.8953	2	4	3	1

Note. C., F., M., T.=The clusters of convective storms, frontal rain, meiyu, typhoons.

doi:10.6342/NTU202001579

A. Table of 297 Stations (Cont'd)

Item	ID	Station	站名	Altitude	Longitude	Latitude	C.	F.	M.	T.
211	C0Z150	Guanghua Ji-An	吉安光華	27	121.595	23.9309	2	4	3	1
212	C0Z160	Fenglin	鳳林	124	121.453	23.7461	2	4	3	1
213	C0Z170	Zhuoxi	卓溪	199	121.303	23.3443	2	4	3	1
214	C0Z180	Sincheng	新城	34	121.605	24.0395	2	4	3	1
215	C1A630	Siapen	下盆	527	121.539	24.7710	4	3	6	5
216	C1AC50	Guandu	關渡	111	121.469	25.1335	8	3	6	4
217	C1C510	Shueiwei	水尾	106	121.087	24.9401	8	6	6	4
218	C1D380	Sinpu	新埔	65	121.038	24.8475	8	6	6	4
219	C1D400	Niaozueishan	鳥嘴山	839	121.284	24.7191	3	6	6	4
220	C1D410	Bailan	白蘭	1290	121.080	24.5794	3	6	6	3
221	C1D420	Taigenan	太閣南	1279	121.161	24.6328	3	6	6	3
222	C1E451	Xiangbi	象鼻	950	120.940	24.3686	3	6	8	3
223	C1E461	Song-An	松安	1325	120.986	24.398	3	7	8	3
224	C1E480	Fongmei	鳳美	576	121.035	24.5548	3	6	6	3
225	C1E511	Xinkai	新開	326	120.827	24.3477	3	6	8	3
226	C1E601	Nanshi	南勢	125	120.735	24.5725	1	6	2	6
227	C1F871	Shangguguan	上谷關	1000	121.019	24.2035	9	7	8	3
228	C1F891	Shaolai	稍來	2205	121.006	24.2635	9	7	8	3
229	C1F911	Xinbogong	新伯公	417	120.841	24.2214	3	6	8	3
230	C1F941	Xueling	雪嶺	2620	121.027	24.2806	9	7	8	3
231	C1G691	Xiashuipu	下水埔	110	120.568	23.8161	1	1	2	6
232	C1H000	Cuifeng	翠峰	2454	121.205	24.1094	9	7	5	2
233	C1H860	Ruiyan	瑞岩	1840	121.184	24.1238	9	7	5	3
234	C1H900	Qingliu	清流	934	120.963	24.0808	7	6	8	3
235	C1H920	Changfeng	長豐	736	120.880	24.1020	7	6	8	3
236	C1H941	Shuangdong	雙冬	630	120.802	23.9674	7	6	8	6
237	C1H971	Liufenliao	六分寮	428	120.638	23.9251	3	1	2	6
238	C1H9B1	Amei	阿眉	1710	120.995	24.1263	7	6	8	3
239	C1I020	Wanda	萬大	1120	121.132	23.9798	7	6	8	3
240	C1I030	Wujie	武界	948	121.053	23.9132	7	6	8	3

Note. C., F., M., T.=The clusters of convective storms, frontal rain, meiyu, typhoons.

doi:10.6342/NTU202001579

A. Table of 297 Stations (Cont'd)

Item	ID	Station	站名	Altitude	Longitude	Latitude	C.	F.	M.	T.
241	C1I070	Heshe	和社	825	120.889	23.5911	7	2	8	3
242	C1I101	Xitou	溪頭	1810	120.808	23.6618	7	2	8	3
243	C1I121	Da-An	大鞍	1515	120.760	23.6784	7	2	8	3
244	C1I131	Tongtou	桶頭	311	120.654	23.6419	5	1	8	6
245	C1I140	Kanaituowan	卡奈托灣	1700	121.088	23.7544	7	2	8	3
246	C1I150	Qingyun	青雲	393	120.949	23.7934	7	6	8	3
247	C1I400	Lingxiao	凌霄	1399	121.005	24.0188	7	6	8	3
248	C1I430	Cuihua	翠華	2415	121.224	24.192	9	7	5	2
249	C1I440	Xingaokou	新高口	2540	120.879	23.4787	10	2	5	2
250	C1I450	Wangxiangshan	望鄉山	3025	120.944	23.5942	10	2	5	2
251	C1I500	Dajianshan	大尖山	2017	120.995	23.8586	7	6	8	3
252	C1I510	Xianjinlindao	線浸林道	1208	120.833	23.7615	7	2	8	3
253	C1M390	Longmei	龍美	1090	120.654	23.4067	5	2	7	8
254	C1M400	Caiguaping	菜瓜坪	369	120.576	23.2519	5	1	7	7
255	C1M480	Dulishan	獨立山	798	120.608	23.5370	5	1	8	8
256	C1N001	Shalun	沙崙	24	120.309	22.9355	6	5	4	7
257	C1O850	Huanhu	環湖	44	120.419	23.1486	6	1	4	7
258	C1O870	Dadongshan	大棟山	1249	120.522	23.3116	5	1	7	3
259	C1O880	Guanshan	關山	223	120.594	23.1734	5	1	7	7
260	C1O921	Nanxi	楠西	115	120.484	23.1835	5	1	4	7
261	C1R110	Gusia	口社	110	120.645	22.7701	5	5	7	7
262	C1R120	Shangdewun	上德文	820	120.704	22.7633	5	5	7	8
263	C1R250	Lili	力里	92	120.629	22.4281	6	5	4	7
264	C1R290	Shihmenshan	石門山	260	120.757	22.1126	6	5	4	7
265	C1S670	Motian	摩天	1580	121.027	23.1995	2	2	3	1
266	C1S880	Shouka	壽卡	474	120.859	22.2389	6	5	4	7
267	C1T800	Luoshao	洛韶	1260	121.454	24.2046	9	4	1	3
268	C1T810	Cih-En	慈恩	2049	121.388	24.1920	9	7	1	2
269	C1T830	Buluowan	布洛灣	675	121.571	24.1718	2	4	3	1
270	C1T920	Zhongxing	中興	68	121.499	23.7695	2	4	3	1

Note. C., F., M., T.=The clusters of convective storms, frontal rain, meiyu, typhoons.

doi:10.6342/NTU202001579

A. Table of 297 Stations (Cont'd)

Item	ID	Station	站名	Altitude	Longitude	Latitude	C.	F.	M.	T.
271	C1T940	Daguan	大觀	539	121.373	23.7142	2	4	3	1
272	C1T950	Tai-An	太安	1050	121.370	23.6667	2	4	3	1
273	C1T970	Danong	大農	183	121.413	23.6152	2	4	3	1
274	C1T980	Longjian	龍澗	1306	121.411	24.0233	2	4	3	1
275	C1T990	Gaoliao	高寮	128	121.357	23.3942	2	4	3	1
276	C1U501	Nioudou	牛鬥	280	121.574	24.6378	4	3	6	5
277	C1U670	Hansi	寒溪	105	121.717	24.634	4	3	6	5
278	C1U690	Sinliao	新寮	101	121.751	24.6256	4	3	6	5
279	C1U880	Beiguan	北關	8	121.872	24.9065	4	3	6	5
280	C1U920	Siyuan	思源	2085	121.347	24.3931	9	7	1	3
281	C1V160	Dakanuwa	達卡努瓦	1040	120.705	23.2798	5	2	7	8
282	C1V170	Paiyun	排雲	3690	120.954	23.4636	10	2	5	2
283	C1V190	Nantianchi	南天池	2700	120.912	23.274	10	2	5	8
284	C1V200	Meishan	梅山	870	120.824	23.2684	10	2	7	8
285	C1V220	Xiaoguanshan	小關山	1781	120.814	23.1542	10	2	7	8
286	C1V231	Gaozhong	高中	731	120.717	23.1349	5	2	7	8
287	C1V300	Yuyoushan	御油山	1637	120.715	23.002	5	2	7	8
288	C1V340	Dajin	大津	190	120.646	22.8883	5	5	7	7
289	C1V390	Jianshan	尖山	60	120.368	22.8132	6	5	4	7
290	C1V570	Jiadong	吉東	82	120.545	22.8542	5	5	4	7
291	C1V580	Xinan	溪南(特生中心)	1656	120.789	23.085	10	2	7	8
292	C1V590	Xinfa	新發	741	120.646	23.057	5	2	7	8
293	C1X040	Dongyuan	東原	232	120.464	23.2916	5	1	4	7
294	C1Z030	Hongye	紅葉	218	121.339	23.4931	2	4	3	1
295	C1Z040	Lishan	立山	434	121.327	23.4434	2	4	3	1
296	C1Z120	Shoufeng	壽豐	62	121.508	23.8709	2	4	3	1
297	C1Z130	Tongmen	銅門	187	121.493	23.9657	2	4	3	1

Note. C., F., M., T.=The clusters of convective storms, frontal rain, meiyu, typhoons.

doi:10.6342/NTU202001579

B. Table of Warning Typhoons from 1998 to 2019

Item	Alert Start Date	Alert End Date	Typhoon	Typhoon (in Chinese)
1	07/08/1998	07/10/1998	NICHOLE	妮蔻兒
2	08/03/1998	08/05/1998	OTTO	奧托
3	09/27/1998	09/29/1998	YANNI	楊妮
4	10/14/1998	10/16/1998	ZEB	瑞伯
5	10/25/1998	10/27/1998	BABS	芭比絲
6	06/05/1999	06/06/1999	MAGGIE	瑪姬
7	08/05/1999	08/08/1999	RACHEL	瑞琪兒
8	08/21/1999	08/21/1999	SAM	山姆
9	10/06/1999	10/10/1999	DAN	丹恩
10	07/05/2000	07/09/2000	KAI-TAK	啟德
11	08/21/2000	08/23/2000	BILIS	碧利斯
12	08/27/2000	08/30/2000	PRAPIROON	巴比侖
13	09/08/2000	09/10/2000	BOPHA	寶發
14	10/23/2000	10/28/2000	YAGI	雅吉
15	10/30/2000	11/01/2000	XANGSANE	象神
16	05/12/2001	05/14/2001	CIMARON	西馬隆
17	06/22/2001	06/23/2001	CHEBI	奇比
18	07/04/2001	07/05/2001	UTOR	尤特
19	07/10/2001	07/11/2001	TRAMI	潭美
20	07/23/2001	07/24/2001	YUTU	玉兔
21	07/28/2001	07/31/2001	TORAJI	桃芝
22	09/06/2001	09/20/2001	NARI	納莉
23	09/22/2001	09/28/2001	LEKIMA	利奇馬
24	10/14/2001	10/17/2001	HAIYAN	海燕
25	07/02/2002	07/04/2002	RAMMASUN	雷馬遜
26	07/08/2002	07/13/2002	NAKRI	納克莉
27	09/04/2002	09/08/2002	SINLAKU	辛樂克
28	04/21/2003	04/24/2003	KUJIRA	柯吉拉
29	06/01/2003	06/03/2003	NANGKA	南卡
30	06/16/2003	06/18/2003	SOUDELOR	蘇迪勒

doi:10.6342/NTU202001579

B. Table of Warning Typhoons from 1998 to 2019 (Cont'd)

Item	Alert Start Date	Alert End Date	Typhoon	Typhoon (in Chinese)
31	08/02/2003	08/04/2003	MORAKOT	莫拉克
32	08/19/2003	08/20/2003	VAMCO	梵高
33	08/31/2003	09/01/2003	DUJUAN	杜鵑
34	11/02/2003	11/03/2003	MELOR	米勒
35	06/08/2004	06/09/2004	CONSON	康森
36	06/30/2004	07/02/2004	MINDULLE	敏督利
37	07/09/2004	07/15/2004	KOMPASU	康伯斯
38	08/11/2004	08/12/2004	RANANIM	蘭寧
39	08/23/2004	08/25/2004	AERE	艾利
40	09/11/2004	09/12/2004	HAIMA	海馬
41	10/24/2004	10/25/2004	NOCK-TEN	納坦
42	12/03/2004	12/04/2004	NANMADOL	南瑪都
43	07/16/2005	07/20/2005	HAITANG	海棠
44	08/02/2005	08/05/2005	MATSA	馬莎
45	08/12/2005	08/13/2005	SANVU	珊瑚
46	08/30/2005	09/01/2005	TALIM	泰利
47	09/09/2005	09/11/2005	KHANUN	卡努
48	09/21/2005	09/23/2005	DAMREY	丹瑞
49	09/30/2005	10/03/2005	LONGWANG	龍王
50	05/16/2006	05/18/2006	CHANCHU	珍珠
51	07/11/2006	07/14/2006	BILIS	碧利斯
52	07/23/2006	07/25/2006	KAEMI	凱米
53	08/06/2006	08/09/2006	BOPHA	寶發
54	08/08/2006	08/11/2006	SAOMAI	桑美
55	09/13/2006	09/16/2006	SHANSHAN	珊珊
56	08/07/2007	08/08/2007	PABUK	帕布
57	08/08/2007	08/09/2007	WUTIP	梧提
58	08/16/2007	08/19/2007	SEPAT	聖帕
59	09/17/2007	09/19/2007	WIPHA	韋帕
60	10/04/2007	10/08/2007	KROSA	柯羅莎

doi:10.6342/NTU202001579

B. Table of Warning Typhoons from 1998 to 2019 (Cont'd)

Item	Alert Start Date	Alert End Date	Typhoon	Typhoon (in Chinese)
61	11/26/2007	11/27/2007	MITAG	米塔
62	07/16/2008	07/18/2008	KALMAEGI	卡玫基
63	07/26/2008	07/29/2008	FUNG-WONG	鳳凰
64	08/20/2008	08/21/2008	NURI	如麗
65	09/09/2008	09/17/2008	SINLAKU	辛樂克
66	09/22/2008	09/23/2008	HAGUPIT	哈格比
67	09/26/2008	09/29/2008	JANGMI	薔蜜
68	06/19/2009	06/22/2009	LINFA	蓮花
69	07/16/2009	07/18/2009	MOLAVE	莫拉菲
70	08/05/2009	08/10/2009	MORAKOT	莫拉克
71	10/03/2009	10/06/2009	PARMA	芭瑪
72	08/30/2010	08/31/2010	NAMTHEUN	南修
73	08/31/2010	09/02/2010	LIONROCK	萊羅克
74	09/09/2010	09/10/2010	MERANTI	莫蘭蒂
75	09/17/2010	09/20/2010	FANAPI	凡那比
76	10/21/2010	10/23/2010	MEGI	梅姬
77	05/09/2011	05/10/2011	AERE	艾利
78	05/27/2011	05/28/2011	SONGDA	桑達
79	06/23/2011	06/25/2011	MEARI	米雷
80	08/04/2011	08/06/2011	MUIFA	梅花
81	08/27/2011	08/31/2011	NANMADOL	南瑪都
82	06/19/2012	06/21/2012	TALIM	泰利
83	06/28/2012	06/29/2012	DOKSURI	杜蘇芮
84	07/31/2012	08/03/2012	SAOLA	蘇拉
85	08/06/2012	08/07/2012	HAIKUI	海葵
86	08/14/2012	08/15/2012	KAI-TAK	啟德
87	08/21/2012	08/25/2012	TEMBIN1	天秤
88	08/26/2012	08/28/2012	TEMBIN2	天秤
89	09/27/2012	09/28/2012	JELAWAT	拉華
90	07/11/2013	07/13/2013	SOULIK	蘇力

doi:10.6342/NTU202001579

B. Table of Warning Typhoons from 1998 to 2019 (Cont'd)

Item	Alert Start Date	Alert End Date	Typhoon	Typhoon (in Chinese)
91	07/17/2013	07/18/2013	CIMARON	西馬隆
92	08/20/2013	08/22/2013	TRAMI	潭美
93	08/27/2013	08/29/2013	KONG-REY	康芮
94	09/19/2013	09/22/2013	USAGI	天兔
95	10/04/2013	10/07/2013	FITOW	菲特
96	06/14/2014	06/15/2014	HAGIBIS	哈吉貝
97	07/21/2014	07/23/2014	MATMO	麥德姆
98	09/19/2014	09/22/2014	FUNG-WONG	鳳凰
99	05/10/2015	05/11/2015	NOUL	紅霞
100	07/09/2015	07/11/2015	CHAN-HOM	昌鴻
101	07/06/2015	07/09/2015	LINFA	蓮花
102	08/06/2015	08/09/2015	SOUDELOR	蘇迪勒
103	08/20/2015	08/23/2015	GONI	天鵝
104	09/27/2015	09/29/2015	DUJUAN	杜鵑
105	07/06/2016	07/09/2016	NEPARTAK	尼伯特
106	09/12/2016	09/15/2016	MERANTI	莫蘭蒂
107	09/15/2016	09/18/2016	MALAKAS	馬勒卡
108	09/25/2016	09/28/2016	MEGI	梅姬
109	10/05/2016	10/06/2016	AERE	艾利
110	07/28/2017	07/30/2017	NESAT	尼莎
111	07/29/2017	07/31/2017	HAITANG	海棠
112	08/20/2017	08/22/2017	HATO	天鴿
113	09/06/2017	09/07/2017	GUCHOL	谷超
114	09/12/2017	09/14/2017	TALIM	泰利
115	07/09/2018	07/11/2018	MARIA	瑪莉亞
116	09/14/2018	09/15/2018	MANGKHUT	山竹
117	07/16/2019	07/18/2019	DANAS	丹娜絲
118	08/07/2019	08/10/2019	LEKIMA	利奇馬
119	08/23/2019	08/25/2019	BAILU	白鹿
120	09/29/2019	10/01/2019	MITAG	米塔

doi:10.6342/NTU202001579

C. Identified Hourly Rainfall Anomalies Caused by Malfunction

Item	Station ID	Station	Time	HR
1	C0A550	Taiping	2001/03/08 04:00	49
2	C0A550	Taiping	2000/06/14 02:00	87.5
3	C0A570	Tonghou	2016/09/08 17:00	62
4	C0A640	Shihding	1998/09/06 14:00	61.5
5	C0A860	Daping	2000/02/15 14:00	73.5
6	C0A880	Fulong	2001/06/07 13:00	46
7	C0A890	Shuangsi	2016/01/25 14:00	74.5
8	C0A890	Shuangsi	2001/06/07 14:00	85.5
9	C0A920	Fugueijiao	2001/03/29 12:00	120
10	C0A920	Fugueijiao	2015/03/23 16:00	61
11	C0A931	Sanhe	1998/12/24 03:00	88.5
12	C0A931	Sanhe	1999/03/23 17:00	116
13	C0A931	Sanhe	2008/11/09 05:00	936
14	C0A940	Jinshan	1999/03/23 16:00	54
15	C0A940	Jinshan	2008/12/24 19:00	735.5
16	C0A940	Jinshan	1998/10/27 12:00	54.5
17	C0A9A0	Dazhi	2007/08/07 15:00	83
18	C0A9A0	Dazhi	2007/08/31 13:00	88
19	C0A9F0	Neihu	2007/08/31 13:00	95
20	C0A9F0	Neihu	2008/07/07 13:00	145
21	C0AH10	Yonghe	1999/06/17 18:00	65
22	C0AH10	Yonghe	1999/06/21 11:00	66
23	C0AH40	Pingdeng	1999/06/17 18:00	81
24	C0AI10	Cyuchih	1998/05/18 16:00	67
25	C0AI10	Cyuchih	2000/06/19 13:00	147
26	C0AI10	Cyuchih	2000/07/10 14:00	119.5
27	C0AI40	Shipai	2008/10/21 14:00	283
28	C0C490	Bade	1998/04/14 06:00	58.5
29	C0C490	Bade	1999/03/23 17:00	62.5
30	C0C490	Bade	1999/12/20 02:00	63

Note.
HR = Hourly Rainfall (mm);
The instrument failed to return the observation in time; the value had accumulated for several hours.

doi:10.6342/NTU202001579

C. Identified Hourly Rainfall Anomalies Caused by Malfunction (Cont'd)

Item	Station ID	Station	Time	HR
31	C0C490	Bade	1998/07/14 15:00	388.5
32	C0C490	Bade	1999/08/05 18:00	75
33	C0C700	Zhongli (NCU)	1998/04/13 16:00	54.5
34	C0D430	Emei	2001/06/25 16:00	77
35	C0D430	Emei	2007/05/07 11:00	60
36	C0D430	Emei	1998/10/07 09:00	136.5
37	C0E430	Nanzhuang	2007/05/07 12:00	101.5
38	C0E520	Dahu	1999/10/06 16:00	60.5
39	C0E790	Zhuolan	1999/08/23 15:00	75.5
40	C0E850	Dahe	2007/05/07 11:00	82
41	C0F970	Dakeng	1999/07/31 20:00	90.5
42	C0G650	Yuanlin	1998/04/15 23:00	67
43	C0G660	Xihu	1998/04/16 11:00	82
44	C0G720	Xizhou	2013/08/29 16:00	105
45	C0H960	Caotun	2003/06/08 08:00	84.5
46	C0H990	Kunyang	2002/07/09 17:00	59.5
47	C0H9C0	Hehuan Mountain	1998/02/20 12:00	137
48	C0H9C0	Hehuan Mountain	1998/02/23 13:00	69
49	C0I110	Zhushan	1998/02/22 03:00	165
50	C0I110	Zhushan	2000/11/09 11:00	90.5
51	C0I380	Jiji	2002/07/09 17:00	153.5
52	C0K240	Caoling	1998/02/22 06:00	128.5
53	C0K240	Caoling	2000/11/09 14:00	152
54	C0K280	Sihu	1998/06/04 04:00	119.5
55	C0K291	Yiwu	1998/06/04 05:00	81.5
56	C0K330	Huwei	1998/06/03 23:00	75
57	C0K330	Huwei	2013/08/29 16:00	145
58	C0K390	Tuku	1998/06/04 05:00	79
59	C0K400	Douliu	2013/08/29 16:00	76
60	C0K410	Beigang	1998/06/04 05:00	63.5

Note.

HR = Hourly Rainfall (mm);

The instrument failed to return the observation in time; the value had accumulated for several hours.

doi:10.6342/NTU202001579

C. Identified Hourly Rainfall Anomalies Caused by Malfunction (Cont'd)

Item	Station ID	Station	Time	HR
61	C0K430	Baozhong	1998/06/04 04:00	75
62	C0K430	Baozhong	2001/07/31 10:00	160
63	C0M410	Matoushan	2005/05/13 13:00	91
64	C0M520	Donghouliao	2001/07/31 08:00	74
65	C0M640	Zhongpu	2001/09/19 10:00	741.5
66	C0M650	Puzi	1998/06/04 05:00	92.5
67	C0M700	Zhuqi	1998/08/07 17:00	79.5
68	C0M710	Dongshi	1998/06/04 04:00	101.5
69	C0O810	Cengwen	1998/02/21 12:00	127.5
70	C0O960	Qiding	2007/05/20 06:00	64
71	C0O970	Hutoupi	2004/10/21 13:00	201
72	C0R130	Ali	2000/04/18 16:00	167.5
73	C0R130	Ali	2001/05/21 11:00	544
74	C0R130	Ali	2001/05/30 11:00	158
75	C0R130	Ali	2001/05/31 11:00	146.5
76	C0R130	Ali	2001/06/14 10:00	86
77	C0R140	Majia	2001/05/21 11:00	828
78	C0R140	Majia	2001/05/31 10:00	434
79	C0R190	Chishan	2003/09/23 08:00	98
80	C0R220	Chaojhou	2016/07/08 10:00	78
81	C0R240	Laiyi	2016/07/08 10:00	114
82	C0R260	Chunri	1999/07/05 11:00	171.5
83	C0R260	Chunri	2016/07/08 10:00	107.5
84	C0R280	Binlang	2012/08/28 15:00	666
85	C0R341	Mudan	2011/09/03 13:00	320.5
86	C0R341	Mudan	2012/07/22 16:00	83
87	C0R341	Mudan	2012/08/25 04:00	460.5
88	C0R341	Mudan	2012/08/28 06:00	163.5
89	C0R580	Nanzhou	2000/11/01 11:00	79
90	C0R580	Nanzhou	2016/07/08 10:00	110

Note.

HR = Hourly Rainfall (mm);

The instrument failed to return the observation in time; the value had accumulated for several hours.

doi:10.6342/NTU202001579

C. Identified Hourly Rainfall Anomalies Caused by Malfunction (Cont'd)

Item	Station ID	Station	Time	HR
91	C0S660	Siama	2016/07/09 14:00	369
92	C0S680	Hongyeshan	1999/08/30 07:00	84
93	C0S680	Hongyeshan	2012/09/26 14:00	158.5
94	C0S680	Hongyeshan	2016/07/09 14:00	157
95	C0S690	Taimali	2012/09/25 17:00	45.5
96	C0S710	Luye	2016/07/09 15:00	226.5
97	C0S760	Hongshih	2000/06/14 10:00	51.5
98	C0S760	Hongshih	1999/09/04 19:00	216.5
99	C0S760	Hongshih	1999/09/14 10:00	49.5
100	C0S760	Hongshih	1999/10/09 11:00	237
101	C0S760	Hongshih	2016/07/09 14:00	176
102	C0S810	Donghe	2005/08/13 08:00	71
103	C0S830	Changbin	2001/07/25 04:00	56.5
104	C0T960	Guangfu	1999/10/13 10:00	131.5
105	C0U720	Nanshan	2002/05/17 12:00	43
106	C0U860	Toucheng	2015/08/17 11:00	182
107	C0U890	Sansing	1999/09/03 09:00	76.5
108	C0U890	Sansing	1999/09/04 21:00	126.5
109	C0U900	Neicheng	2000/11/13 16:00	166.5
110	C0V250	Jiasian	2005/05/13 14:00	82
111	C0V310	Meinong	2000/04/18 16:00	118
112	C0V310	Meinong	2001/05/21 10:00	312.5
113	C0V310	Meinong	2001/05/30 11:00	131
114	C0V310	Meinong	2001/05/31 10:00	113
115	C0V310	Meinong	2001/06/14 10:00	86
116	C0V350	Xipu	2001/05/21 10:00	342
117	C0V350	Xipu	2001/05/30 11:00	166
118	C0V350	Xipu	2001/05/31 09:00	101
119	C0V370	Gutingkeng	1999/08/12 12:00	70.5
120	C0V400	Agongdian	1999/08/12 12:00	67

Note.

HR = Hourly Rainfall (mm);

The instrument failed to return the observation in time; the value had accumulated for several hours.

doi:10.6342/NTU202001579

C. Identified Hourly Rainfall Anomalies Caused by Malfunction (Cont'd)

Item	Station ID	Station	Time	HR
121	C0V440	Fengshan	2016/07/08 10:00	96
122	C0V450	Fengsen	2016/07/08 10:00	122.5
123	C0V730	Daliao	2000/08/11 14:00	83
124	C0V730	Daliao	2016/07/08 10:00	123
125	C0V740	Qishan	2000/04/19 13:00	114
126	C0V740	Qishan	2001/05/21 11:00	292
127	C0V740	Qishan	2001/05/30 11:00	135.5
128	C0V740	Qishan	2001/05/31 10:00	88.5
129	C0V740	Qishan	2001/06/14 10:00	93.5
130	C0V750	Luzhu	1999/08/12 12:00	67
131	C0V770	Dashe	1999/07/05 12:00	79
132	C0X050	Donghe	2001/07/31 07:00	148.5
133	C0X080	Jiali	2016/09/03 11:00	93
134	C0X080	Jiali	2016/09/07 13:00	135
135	C0X210	Baihe	2001/07/31 08:00	167.5
136	C0X230	Yanshui	2001/07/31 07:00	93
137	C0X250	Xinying	1998/09/09 16:00	79.5
138	C0X250	Xinying	2001/07/31 08:00	116
139	C0X290	Beimen	1998/06/04 04:00	68.5
140	C1A630	Siapen	2000/01/31 14:00	59.5
141	C1A630	Siapen	2001/07/17 13:00	56
142	C1A630	Siapen	2001/07/21 11:00	130.5
143	C1A630	Siapen	2001/07/21 14:00	216
144	C1A630	Siapen	2001/08/19 17:00	195
145	C1A630	Siapen	2001/08/24 19:00	94
146	C1AC50	Guandu	2001/08/29 21:00	61
147	C1C510	Shueiwei	1999/03/23 17:00	50
148	C1C510	Shueiwei	2015/06/23 23:00	55
149	C1D380	Sinpu	1999/03/23 12:00	67
150	C1D400	Niaozueishan	1998/09/30 08:00	113

Note.

HR = Hourly Rainfall (mm);

The instrument failed to return the observation in time; the value had accumulated for several hours.

doi:10.6342/NTU202001579

C. Identified Hourly Rainfall Anomalies Caused by Malfunction (Cont'd)

Item	Station ID	Station	Time	HR
151	C1D400	Niaozueishan	1999/08/05 18:00	67.5
152	C1D420	Taigenan	1999/06/15 16:00	79
153	C1D420	Taigenan	1998/09/30 11:00	133
154	C1E451	Xiangbi	1998/02/20 18:00	135
155	C1E461	Song-An	1998/02/20 13:00	163
156	C1E480	Fongmei	1998/02/25 18:00	268
157	C1F871	Shangguguan	1998/02/20 15:00	196.5
158	C1F891	Shaolai	1998/02/20 17:00	223.5
159	C1F911	Xinbogong	1998/07/08 16:00	90
160	C1F941	Xueling	1998/02/20 18:00	251.5
161	C1F941	Xueling	2006/06/02 21:00	76.5
162	C1H000	Cuifeng	2002/07/09 17:00	39.5
163	C1H860	Ruiyan	2002/07/09 17:00	59.5
164	C1H920	Changfeng	2005/08/08 13:00	186.5
165	C1H941	Shuangdong	1998/04/15 22:00	65
166	C1H9B1	Amei	1998/02/20 14:00	149.5
167	C1H9B1	Amei	2006/05/03 15:00	113.5
168	C1H9B1	Amei	2006/06/02 21:00	65.5
169	C1I101	Xitou	1998/04/15 21:00	47
170	C1I121	Da-An	1998/02/21 19:00	98.5
171	C1I131	Tongtou	1998/02/21 19:00	109.5
172	C1I150	Qingyun	1998/01/16 18:00	75
173	C1I440	Xingaokou	2006/06/10 15:00	55.5
174	C1I440	Xingaokou	2011/05/15 12:00	65
175	C1I440	Xingaokou	2011/05/18 12:00	160.5
176	C1I450	Wangxiangshan	2002/07/09 17:00	65
177	C1I510	Xianjinlindao	1998/01/16 18:00	68
178	C1O880	Guanshan	2005/05/13 13:00	110
179	C1R110	Gusia	2001/05/21 14:00	579.5
180	C1R110	Gusia	2001/05/23 09:00	97

Note.

HR = Hourly Rainfall (mm);

The instrument failed to return the observation in time; the value had accumulated for several hours.

doi:10.6342/NTU202001579

C. Identified Hourly Rainfall Anomalies Caused by Malfunction (Cont'd)

Item	Station ID	Station	Time	HR
181	C1R110	Gusia	2001/05/29 09:00	87
182	C1R110	Gusia	2001/05/31 16:00	334
183	C1R110	Gusia	2001/06/14 09:00	82
184	C1R110	Gusia	1999/08/12 12:00	82
185	C1R120	Shangdewun	2001/05/21 11:00	711
186	C1R120	Shangdewun	2001/05/30 16:00	198
187	C1R120	Shangdewun	2001/05/31 10:00	126
188	C1R120	Shangdewun	2001/06/14 10:00	94
189	C1R250	Lili	1999/07/05 12:00	115.5
190	C1R250	Lili	2016/07/08 10:00	116.5
191	C1S670	Motian	2016/07/09 13:00	216.5
192	C1S880	Shouka	2005/07/20 15:00	148.5
193	C1S880	Shouka	2011/09/03 14:00	99.5
194	C1T800	Luoshao	2005/07/18 10:00	98
195	C1T970	Danong	2009/09/27 05:00	69
196	C1U690	Sinliao	2009/10/12 14:00	734.5
197	C1V220	Xiaoguanshan	1999/11/24 15:00	99.5
198	C1V220	Xiaoguanshan	2005/05/13 13:00	130
199	C1V300	Yuyoushan	1999/07/05 12:00	83.5
200	C1V340	Dajin	1998/10/07 15:00	77.5
201	C1V580	Xinan	2005/05/13 13:00	106
202	C1V580	Xinan	2006/06/08 17:00	76.5
203	C1V590	Xinfa	2003/10/13 21:00	137.5
204	C1Z130	Tongmen	2005/09/22 10:00	118.5
205	C1Z130	Tongmen	2005/09/23 09:00	364.5

Note.

HR = Hourly Rainfall (mm);

The instrument failed to return the observation in time; the value had accumulated for several hours.

doi:10.6342/NTU202001579