Chemical composition of rainwater harvested in East Malaysia

Siti Nor Fazillah Abdullah¹, Azimah Ismail¹,²†, Hafizan Juahir¹, R Badlishah Ahmad³, Fathurrahman Lananan¹, Noor Muzlinda Hashim¹,⁴, Nadiana Ariffin¹,⁵, Munirah Abdul Zali⁵, Tengku Azman Tengku Mohd⁵, Md Hafiz Fikrie Hussin⁵, Raja Intan Sariah Raja Mahmood¹,⁵, Jef Rizal Abdul Jamil¹, Safari Mat Desa⁸

¹East Coast Environmental Research Institute, Universiti Sultan Zainal Abidin, Gong Badak Campus, 21300 Terengganu, Malaysia
²Faculty of Innovative Design & Technology, Universiti Sultan Zainal Abidin, Gong Badak Campus, 21300 Terengganu, Malaysia
³School of Computer & Communication Engineering, c/o School of Manufacturing Engineering Complex, Universiti Malaysia Perlis, Pauh Putra Campus, 02600 Aru, Perlis
⁴Muadzam Shah Polytechnic, Lebuhraya Tun Razak, 26700, Muadzam Shah, Pahang, Malaysia
⁵Kuala Terengganu Polytechnic, Jalan Sultan Ismail, 20200, Kuala Terengganu, Terengganu, Malaysia
⁶Environmental Health Division, Department of Chemistry Malaysia, Ministry of Science, Technology and Innovation, Jalan Sultan, Petaling Jaya, Selangor, Malaysia
⁷Integrated Envirotech Sdn. Bhd., Lot 32-2, Jalan Setiawangsa 11A, Taman Setiawangsa, Kuala Lumpur, Malaysia
⁸River Basin Research Center, National Hydraulic Research Institute of Malaysia (NAHRIM), Jalan Putra Permai, 43300, Seri Kembangan, Malaysia

Abstract

As part of the implementation of a rainwater harvesting system as an alternative water source supply for non-potable use, therefore the characteristic of chemical compounds was significantly explored. The Department of Chemistry, Malaysia, gave the data set for three years (2017-2019). Some chemometric techniques, including PCA, were performed to identify the dimensionality of the rainwater data, hence establishing the rainfall index’s purity to determine the quality of rainwater in the study area. Discriminant analysis managed to differentiate each rain gauge station. Cluster analysis was then applied to perform smaller group of rain gauge stations. The result demonstrates that sea salt, secondary aerosols, trace metals, crustal origin, and organic acid dominated the dimensionality of rainwater data with a total variance of 53.38% and indicated that the PRI was significantly diversified into good purity of rainfall index (GPRI), (Labuan and Danum Valley), moderate purity of rainfall index (MPRI), (Kuching and Tawau) and bad purity of rainfall index (BPRI), (Kota Kinabalu and Bintulu). From the study, it can be stipulated that the chemical composition of rainwater in the study area was attributable to the local activities.

Keywords: Chemical composition, Chemometric, non-potable, Principal component analysis, Rainwater harvesting
1. Background

1.1. Preliminary

Rainwater has been confirmed polluting in most regions of the world due to numerous pollutants loads in the atmosphere [1]. Understanding the chemical composition of rainwater is important to enable the investigation the atmospheric conditions of a region and the concentration of the soluble components that contribute to rainwater chemistry [2, 3]. The characteristic of rainwater depends on the atmospheric particulate or gaseous constituents produced locally or transported from distant sources by natural or anthropogenic sources [4]. Owing to the effect of local sources, the chemical composition of rainwater varies by geographical locations [5]. Emission of air contaminants is tremendously increased primarily from the anthropogenic activities in urban area [6, 7] as resulted from large populations, fast-growing economics, high energy use, agricultural development, and industrialization [6]. A study was done in a lofty industrial activity like New Castle, the harvested rainwater was found to be seriously contaminated mostly by trace metals [8]. Contrary to another study by Al-Khashman [3], which was carried out in Jordan, the concentration of trace metals was relatively low due to the local condition influenced by local anthropogenic sources.

The scavenging of air contaminants influences both the chemical composition and the pH of rainwater [9]. Acid precipitation is mostly caused by the anthropogenic sources such as sulfur oxides, and nitrogen oxides [6, 10]. However, the acidity of rainwater depends on the relative contribution and neutralization of the major acidic and alkaline ions in atmosphere [3]. It is important to determine the characteristics of ions in water, both ionic and cationic. Thus, both
ionic and cationic has become the focus of extensive study since the last two decades because of environmental concerns [1, 9].

1.2. The Quality of Rainwater Harvesting

Historically, rainwater harvesting has been practiced traditionally in many dry regions in the world as early as 4,500 B.C [10, 11]. Currently, rainwater harvesting has been practiced in many countries for reducing reliance on the availability of domestic water from dams and reservoirs [12], supply non-potable water to building in urban areas [13], and solving water issues in semi-arid regions for agricultural and domestic uses [14]. Yet, some constraints should be pointed out before the harvested rainwater could be used, especially if it is intended to be used for potable uses [15]. Many international studies reported the quality of rainwater is contaminated with various microbial pollutants [16-18] and heavy metals originated from raindrops, catchment areas, and storage [19], which requires proper treatment.

The system of rainwater harvesting is collecting and storing rainwater that falls upon the roof surface for later use [20]. Three principal components are required to harvest the rainwater: namely the catchment area, the collection device, and the conveyance system [21]. Since the collected rainwater can supply the water with various economic and environmental benefits, it can significantly provide water usage for non-potable without requiring a drinking water quality standard [22]. The presence of pollutants might not be of significant concern, and the criteria for treatment may be less strict or not at all necessary [23]. The appropriate treatment of rainwater is needed if it is meant for drinking. The system such as the first flush is used to decrease the accumulation of pollutants [24], a post-cistern treatment proven to reduce the amount of total
coliorm [25] and slowing sand filtration to reduce the turbidity and heavy metal concentration 
[26, 27]. In a broader sense, rainwater harvesting is necessary to reduce urban water 
consumption, increasing its productive use [28]. Furthermore, rainwater harvesting can achieve 
sustainability and water conservation supplies around the world [13].

1.3. East Malaysia Climate and Rainwater Blessing

Malaysia is a developing country with continued economic development and population 
expansions [29, 30]. The rapid growth in Malaysia has led to several environmental problems,
particularly the water stress problem [31, 32]. Therefore, the government has acknowledged the 
water scarcity problem occurs due to the increasing demand for water supply, water sources 
pollution, and encroachment into catchment [33]. With the blessing of abundance rainfall, there 
is a lot of potential that can be explored to meet the water shortage problem in Malaysia. The 
quality of rainwater is assessed in terms of its physical, chemical, and microbiological 
characteristics. The chemical content of harvested rainwater usually adheres to the WHO [34]. 
The rainwater samples were analyzed for total dissolved solids, conductivity, pH, major cations, 
and anions [35]. The composition of anions and cations were determined by ion chromatography, 
while pH and electrical conductivity were determined using pH meter and conductivity meter, 
respectively [36].

Generally, rainwater is relatively free from major pollutants in Malaysia [37]. According 
to Hashim et al. [38], the rainwater quality in Malaysia satisfies the drinking water standard. The 
finding is supported by Asman et al. [39]. Based on their study in Bangi; rainwater harvesting 
can be supplied for domestic purposes and as well as drinking water, but with additional
treatment. Simultaneously, a study done in Sandakan stated that rainwater quality is always better than surface and groundwater quality [40]. However, in an urban area like Kuala Lumpur, rainwater’s pH value is slightly acidic [41, 42]. Thus, the use of rainwater for non-potable purposes is recommended [43].

The Ministry of Housing and Local Government (MHLG) officially launched the rainwater harvesting system in Malaysia after the drought event of 1998 [38]. The initial acceptance of the rainwater harvesting method is not strong enough. Therefore, the Ministry of Energy, Green Technology and Water (KTTHA) had introduced two new water laws: Water Services Industry Act 2006 and Water Services Commission Act 2006, which encourages the implementation of rainwater harvesting system [35]. In 2012, the Malaysian government imposed new rainwater harvesting system at large commercial and residential buildings, such as bungalows and semi-detached houses, to install a rainwater harvesting system under the Uniform Building By-Laws 1984 [44].

Different multivariate analyses have been used in modern research to analyze the chemical composition of water quality [45-48]. To date, limited studies were reported where multivariate statistical techniques were applied in rainwater study. Thus, this study proposed a standard multivariate analysis method, including principal component analysis, discriminant analysis, and cluster analysis to precisely investigate the characteristics of rainwater in the study area. East Malaysia has recently been selected for in depth research to study the composition of rainwater’s chemical characteristics as an alternative water resource, particularly for non-potable uses due to its water shortage problems [40]. Therefore, it is necessary to analyze the characteristics of the rainwater and provide a reliable assessment of rainwater quality for the
consumption of humans. The research however was to investigate the quality of rainwater in the study area for the potential use of non-potable benefits. Besides, it is believed that rainwater harvesting will be one of the solutions to overcome the water shortage problem worldwide.

2. Research Methodology

2.1. Geography and Rain Gauge Stations

East Malaysia comprises of two largest states, Sabah and Sarawak that geographically lies on the northwestern coast of Borneo Island. The state of Sabah is covering the area of 73,856 km$^2$, while the state of Sarawak with the maximum area of 124,989 km$^2$ sharing the border of Kalimantan, Indonesia. On the northern part of Sarawak State has land frontiers with two enclaves which make up Brunei (Fig. 1). East Malaysia experiences a wet and humid tropical climate with rainfall season virtually throughout the year with the average rainfall of 5,080 mm annually [49]. The monthly cumulative distribution of rainfall is influenced by the seasonal monsoons, namely Northeast monsoon (October to March) and the southwest monsoon (April to September) [50]. The northeast monsoon is the primary rainy season in Malaysia, produces heavy rainfall in Sarawak. The southwest monsoon is relatively drier except in Sabah. The rain events consist of convective and widespread rain. Convective rain is characterized by the intense rainfall over a short period and covers a limited area [51]. The quantities of rainfalls are changeable depending on the seasons of the year. There are 504 rain gauge stations scattered around Bintulu, Kota Kinabalu, Kuching, Labuan, Danum Valley, and Tawau.
2.2. Data Collection

In this study, the hourly rainfall of secondary data was sourced from the Department of Chemistry Malaysia. All stations were identified based on data available starting from January 1, 2017, to December 31, 2019. The chemical variables detected in this study were ammonium (NH$_4^+$), calcium (Ca$^{2+}$), fluoride (F$^-$), magnesium (Mg$^+$), potassium (K), sodium (Na$^+$), nitrate (NO$_3^-$), sulfate (SO$_4^{2-}$), acetate (C$_2$H$_3$O$_2^-$), chloride (Cl$^-$), formate (CHO$_2^-$), methane sulfonic acid (CH$_4$O$_3$S), oxalate (C$_2$O$_4^{2-}$), copper (Cu), iron (Fe), manganese (Mn), mercury (Hg), nickel (Ni), cadmium (Cd), conductivity (EC), lead (Pb), pH and zinc (Zn). The focus of this study only on statistical analysis via chemometric approaches. All mathematical and statistical computations were performed using Excel 2013 (Microsoft Office). Principal component analysis (PCA), discriminant analysis (DA), and Hierarchical algorithm cluster analysis (HACA) were performed via XL STAT add-in software.
2.3. Principal Component Analysis

PCA is a powerful supervised pattern recognition technique used to explain the variance in a dataset of inter-correlated variables. It is a technique applied to reduce the original variables into a smaller number called principal components (PCs), accounting for most of the variance in the observed variables [52]. The PC can be expressed as:

\[ z_{ij} = a_{i1}x_{1j} + a_{i2}x_{2j} + \ldots + a_{im}x_{mj} \]  

(1)

Where \( z \) is the component score, \( a \) is the component loading, \( x \) is the measured value of the variable, \( i \) is the component number, \( j \) is the sample number, and \( m \) is the total number of variables. PCA reduces the dimensionality of a data set consisting of a large interrelated variable while remaining as much as possible the variability present in the data set [53]. The number of components to keep was based on the Kaiser criterion, for which only the components with an eigenvalue greater than one are retained [54]. The eigenvalues of the PCs are a measure of their associated variance, the loadings give the participants of the original variables in the PCs, and the individual transformed observations called scores. [55] classified the factor loadings as 'strong,' 'moderate,' and 'weak,' corresponding to absolute loading values of > 0.75, 0.75-0.5, and 0.50-0.30. After applying PCA, varimax rotation based on factor loading was conducted, allowing the 'cleaning up' of the PCs by increasing the participation of the variables with a higher contribution and simultaneously reducing the variables with a lower contribution [56]. In this study, PCA was performed on the dataset having the dimensions of 504 objects and 23 physicochemical and metal parameters to define many variables into a smaller set. Hence, identify the latent factor which contributes to the existing these parameters in the rainfall data.
2.4. Developing Purity of Rainfall Index

Using the value of factor scores generated via PCA, each factor can be viewed as one aspect of rainfall. Therefore, factor scores can be used as a single index indicating the aspect with which the factor associated. PRI was a composite of different variables. It was developed by weighting each factor score with the respective variance using the equation below:

\[ PRI = \sum_{i}^{n} F_i w_i \]  

Where \( n \) is the number of factors selected, \( F_i \) is factor \( i \) score and \( w_i \) is the percentage of the variance factor of \( i \). The weightage (summation of the factor score loadings multiplied with variability) gave the insignificant values (comprising of negative and positive values) in which impossible to achieve the desired objective or targeted index. Subsequently, the normalization or rescaling was performed using the Eq. (3) to \( z \)-values (such that the variance for each variable would equal unity), rescaling the index through weightage to scale the range from 1 to 100 (statistical rules) using the Univariate analysis to obtain the accurate index value.

\[ Rescaling \ (1 \ to \ 100) = a + [(x_i - A)(b - a)/(B - A)] \]  

Where \( a \) equal to 1, \( x_i \) is the actual observation, \( A \) and \( B \) respectively, are the lowest and highest factor score, and \( b \) is the constant value of 100. In this study, univariate clustering was used to execute PRI divided into three categories viz named Good, Moderate, and Bad. The lowest PRI value indicates that the characteristics of the rainwater at the rain monitoring stations were not much affected by the contaminants. Therefore, the areas were compatible to be considered as good rainwater quality.

2.5. Discriminant Analysis
Discriminant analysis is a supervised pattern recognition technique used to determine the variables responsible for separating the observations into different groups [57, 58]. Linear combinations of the independent variables found through this technique will discriminate against the groups so that the misclassification error rates are minimized. Discriminant analysis was performed on original data without affecting the results and comparability with other multivariate methods and constructed a discriminant function (DFs) for each group as follows:

\[ f(G_i) = k_i + \sum_{j=1}^{n} w_{ij} p_{ij} \]  

where \( i \) represents the number of groups (G), \( k_i \) is the constant inherent to each group, \( n \) is the number of parameters used to classify a set of data into a given group and \( w_j \) is the weight coefficient assigned by DF analysis (DFA) to a given parameter (\( p_j \)) [59]. In this study, the purity of the rainfall index executed by PCA was treated as the dependent variable, while the 23 physicochemical and metal parameters were treated as independent variables. DA was performed on the original data via standard, forward stepwise and backward stepwise modes, to construct the best DFs to confirm the indices developed by PCA.

2.6. Hierarchical Algorithm Cluster Analysis

Cluster analysis is an assortment of techniques designed to perform classification by assigning observation to the group, so each is more or less homogeneous and distinct from other groups. Hierarchical algorithm cluster analysis (HACA) is the most common approach and provides intuitive similarity between anyone sample and the entire data set [60-62]. HACA is an unsupervised pattern recognition technique used to identify the natural grouping pattern and group variables without making any prior assumption. The objects with high similarity are
clustered into the same group while the objects with high differences are clustered in different
groups. The result of HACA illustrated using a dendogram through Ward's method using the
squared Euclidean distance. The Ward's method, using squared Euclidean distances as a measure
of similarity, possesses a small space distorting effect, uses more information about cluster
contents than other methods, and has been proved to be an extremely powerful grouping
mechanism [56]. The dendrogram provides a visual summary of the clustering process,
presenting a picture of the groups and their proximity, with a dramatic reduction in the
dimensionality of the original data [63]. In this study, HACA was performed on the rainfall data
set by means of Ward's method, using Euclidean distance.

3. Presentation of Findings and Discussion

3.1. Atmospheric Particulate Matter in Rainwater

Summary statistics for chemical composition in the study area are presented in Table 1. Most
parameters displayed a wide variation in elemental concentration, as reflected by large standard
deviations. Chemical parameters like ammonium, potassium, sodium, nitrate, sulfate, and
chloride had the greatest standard deviation values in the data set, as shown in the table, and
were 39.80, 10.24, 28.00, 21.70, 10.60, and 33.78 mg/L, respectively. The highest variance of
ammonium within the study area could have resulted from the value of ammonium concentration
in each rain gauge station is far from each other which means the concentration at each rain
gauge station very distinctive due to the locations. The highest concentration of variance was due
to the local activities at each rain gauge station. The agricultural areas such as Tawau, was
considerably emitted higher ammonia concentration into the atmosphere than the industrial areas
like Kuching and Bintulu [64]. The emission of ammonia to the atmosphere associated with livestock production [65].

Among the inorganic species, chloride was the most abundant species by mass, followed by sodium, ammonium, nitrate, and sulfate. The high concentration of chloride ion is acquired from the large bodies of the South China Sea. While among the metals, Zn was the most abundant by mass than the other metals with mean value of 0.30 mg/L. This study further suggested the potential factor of zinc emission in the rainwater subjected to the roofing material [13]. Zinc-based roofing material undergoing the chemical reaction with the presence of oxygen in atmosphere to form zinc oxide, and zinc hydroxide if expose to moisture. The pH values of collected rainwater samples ranging from 4.3-7.0, thus it noticeable the rainwater within the study area were acidic. The acidity level increased dramatically as the anions level in the atmosphere increased, thus lower the pH values [66]. However, the optimum value of the pH level in the study area seems to be neutralized in some rain gauge stations by alkaline ions (NH$_4^+$, Ca$^{2+}$, Mg$^+$, Na$^+$) [67].

3.2. Chemical Composition of Rainwater

Based on the chi-square value, it was calculated as 291.10 from Bartlett's sphericity test (d.f. = 253 p < 0.001), and the Kaiser-Meyer-Olkin (KMO) test, the sampling adequacy was greater than 0.5. The output revealed, PCA was significant in dimensional reduction of the complex rainfall dataset, thus subjected for for further analysis [68]. The PCA with varimax rotation explained that out of 23 principal components, only eight PCs with an eigenvalue greater than 1.0 (Table 2), for the total explained variance of 67.62% was considered for further analysis. The
most significant PCs were the first five, with a cumulative explanation of 53.38% of data variability (variance for PC1 = 23.25%, PC2 = 9.39%, PC3 = 8.68%, PC4 = 7.02% and PC5 = 5.04%). The remainder PCs, with a variance of 4.98%, 4.69% and 4.56%, respectively, did not reveal any significant similarities among the rain gauge stations.

The factor loadings provide a correlation between the chemical composition with the factors (Table 3). The varimax factor (VF) loading plot (Fig. 2) revealed the strong correlation and dependence between the chemical composition in rainwater (shorter distance corresponding to a stronger correlation between the parameters) [68]. The first factor accounted for 23.25% of the total variance and had a strong positive correlation with sea salt ions (Mg\(^+\), Na\(^+\) and Cl\(^-\)) and moderate positive loading with heavy metal (Ni). Mg\(^+\), Na\(^+\) and Cl\(^-\) are the major ions in aqueous solution. The strong correlation among these ions explains it came from the same sources such as marine aerosols [69].

In this case, Ni loading (0.51) was not as high as the loadings of the other elements of the group, which may therefore, implied that Ni has independent behavior within the group. Ni is one of the trace metals released from both natural sources and anthropogenic activities [70]. Nickel presents in the air as a result of the industrial activities including combustion of coal, diesel oil, fuel oil, the incineration of waste and sewage, stainless steel production, petrochemical plant, brick manufacture, and heavy ship traffic [71]. Despite nickel moderately loaded on this factor, the presence of this stand-alone metal without any toxic metals loaded in the third factor may explain, the abundance of nickel in the study area expected was emitted from multiple sources. In the study area, Nickel could possibly contributed from natural sources such as vegetation and sea salt [72], while the anthropogenic sources, including both stationary and
mobile sources [73]. Therefore, this factor shows the origins of sea salt in this study derived from
the sea salt spray and industrial activity.

Table 1. Descriptive Statistics of Rainwater Variables in Eastern Malaysian between 2017 and 2019

| Parameter | Unit | Min. | Max. | Median | Mean | Variance | SD |
|-----------|------|------|------|--------|------|----------|----|
| NH₄⁺ | mg/L | 0.05 | 643.58 | 0.97 | 9.64 | 1583.97 | 39.80 |
| Ca²⁺ | mg/L | 0.05 | 64.16 | 2.18 | 3.53 | 24.27 | 4.93 |
| F⁻ | mg/L | 0.03 | 5.37 | 0.10 | 0.33 | 0.29 | 0.54 |
| Mg⁺ | mg/L | 0.10 | 34.26 | 0.99 | 2.08 | 12.44 | 3.53 |
| K | mg/L | 0.04 | 175.53 | 1.40 | 3.41 | 104.83 | 10.24 |
| Na⁺ | mg/L | 0.20 | 314.05 | 9.96 | 18.36 | 784.25 | 28.00 |
| NO₃⁻ | mg/L | 0.01 | 425.51 | 2.42 | 6.16 | 471.05 | 21.70 |
| SO₄²⁻ | mg/L | 0.01 | 187.41 | 3.52 | 5.63 | 112.36 | 10.60 |
| C₂H₃O₂⁻ | mg/L | 0.03 | 32.31 | 0.03 | 0.35 | 4.38 | 2.09 |
| Cl⁻ | mg/L | 0.10 | 355.91 | 12.19 | 22.18 | 1141.28 | 33.78 |
| CHO₂⁻ | mg/L | 0.02 | 6.92 | 0.05 | 0.13 | 0.26 | 0.51 |
| CH₄O₃S | mg/L | 0.02 | 4.42 | 0.02 | 0.05 | 0.07 | 0.27 |
| C₂O₄(2⁻) | mg/L | 0.01 | 1.40 | 0.04 | 0.08 | 0.02 | 0.12 |
| Cu | mg/L | 0.00 | 1.60 | 0.01 | 0.03 | 0.01 | 0.09 |
| Fe | mg/L | 0.04 | 0.73 | 0.04 | 0.07 | 0.01 | 0.08 |
| Mn | mg/L | 0.00 | 1.15 | 0.04 | 0.08 | 0.01 | 0.11 |
|     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|
| Hg  | mg/L| 0.00| 0.00| 0.00| 0.00| 0.00| 0.00|
| Ni  | mg/L| 0.02| 0.68| 0.02| 0.02| 0.00| 0.04|
| Cd  | mg/L| 0.00| 0.01| 0.00| 0.00| 0.00| 0.00|
| EC  | -   | 0.12| 22.20|0.67| 0.95| 1.76| 1.33|
| Pb  | mg/L| 0.00| 0.04| 0.00| 0.00| 0.00| 0.00|
| pH  | -   | 4.25| 7.03| 5.20| 5.27| 0.18| 0.43|
| Zn  | mg/L| 0.00| 6.10| 0.11| 0.30| 0.36| 0.60|

1 Abbreviation for physicochemical and metal compounds in this study: ammonium (NH$_4^+$), calcium (Ca$^{2+}$), fluoride (F$^-$), magnesium (Mg$^+$), potassium (K), sodium (Na$^+$), nitrate (NO$_3^-$), sulfate (SO$_4^{2-}$), acetate (C$_2$H$_3$O$_2$), chloride (Cl$^-$), formate (CHO$_2^-$), methane sulfonic acid (CH$_4$O$_3$S), oxalate (C$_2$O$_4$(2-)), copper (Cu), iron (Fe), manganese (Mn), mercury (Hg), nickel (Ni), cadmium (Cd), conductivity (EC), lead (Pb), pH and zinc (Zn). While SD refers to the standard deviation.

2 **Table 2.** The Eigenvalues and Variability of Percentage Contributed by Eight PCs

|     | PC1  | PC2  | PC3  | PC4  | PC5  | PC6  | PC7  | PC8  |
|-----|------|------|------|------|------|------|------|------|
| Eigenvalue | 5.35 | 2.16 | 1.99 | 1.61 | 1.16 | 1.15 | 1.08 | 1.05 |
| Variability (%) | 23.25 | 9.39 | 8.68 | 7.02 | 5.04 | 4.98 | 4.69 | 4.56 |
| Cumulative (%) | 23.25 | 32.64 | 41.32 | 48.34 | 53.38 | 58.36 | 63.05 | 67.62 |
The second factor consisting of high positive loading on potassium and electrical conductivity, and moderate positive loading on ammonium, nitrate, and sulfate, explained 9.39% of the variance. The average positive loading of three inorganic ions (ammonium, nitrate, and sulfate) suggests that the precursors of these ions were released from similar emission sources such as coal burning, vehicle exhaust or industrial sector besides meteorological conditions and internal mixing [74]. However, ammonium shows higher loadings compared to the nitrate and sulfate. It was possible that ammonium might have resulted from the reaction of other particles in the airborne and played a significant role as a neutralizing agent [75]. Ammonium, sulfate, and nitrate are secondary inorganic aerosol, they were formed by physical or chemical reactions of precursor gases, such as sulfuric acid, nitric acid, and ammonia [76]. Sulfuric and nitric acids are atmospheric oxidation products of gaseous sulfur dioxide and nitrogen oxides, respectively. The strong correlation among nitrate coupled with sulfate, produced as secondary aerosols, originated
from similar sources of major industrial sources, including the power industry, biomass industry, vehicular emission, and residential [77, 78]. However, the modest correlation between potassium with nitrate and sulfate indicating the formation of nitrate and sulfate contributed from biomass burning aerosols [79].

The generation ammonia was probable the contribution from the agricultural practices such as livestock emission, fertilizer use, and agricultural waste burning [80]. The reaction of ammonia and oxygen in the air, it would have undergone for conversion to ammonium aerosol, subjected to the acid concentration in the air. The presence of ammonia in the air led to the production and stabilization of particulate sulfate and nitrate [81]. Enough gaseous ammonia (NH₃) may react with sulfate and nitrate to form ammonium sulfate (NH₄)₂SO₄ and ammonium nitrate (NH₄NO₃) via particle gas formation and gas to particle conversion [76].

Conversely, potassium did not show a significant correlation with these secondary inorganic ions. Though, it was the highest loading in these factors compared to the other three ions. It may imply that this ion contributed from multiple sources such as biomass combustion, biogenic process, sea salt, and soil dust [82, 83]. The prior study reported the presence of potassium in the atmosphere generated from biomass burning [84, 85]. Since EC is strongly associated with the degree of acidity of the rainwater, this factor explains the loading of this parameter. The high concentration of inorganic particles in the rainwater constitutes a higher level of salinity, thus increased the electrical conductivity of the rainwater [55]. Factor two explains the sources of these chemical compounds originated from the industrial activities.

The third factor acknowledged trace metals copper (Cu), cadmium (Cd), lead or plumbum (Pb), and zinc (Zn) with positive loading, 8.68% of the total variance. A significant
positive correlation among these metals, suggesting a possible common origin of these metals came from similar sources. This factor should be emissions from various sources, considered contributed by the vehicular exhaust, industrial activities, oil and waste combustion in incinerator and coal burning in electricity generation plants [86, 87]. Yet, this factor shows there were two parts of toxic trace metals load. Cu, coupled with Pb, has strong loading, while, Zn, coupled with Cd, has moderately positive loading. Cu and Pb were significantly correlated based on the loading value in this factor and maybe originated mainly from industrial sources and traffic pollution [88]. Cu can be emitted from the linings of a vehicle, especially during congested traffic conditions, while Pb comes from the usage of leaded gasoline and batteries [89].

While both Cd and Zn have moderate positive loading in this factor, it was generated from a similar source, the possible sources are predominantly the earth's crust, as well as the anthropogenic sources [90]. The primary anthropogenic sources for these Cd and Zn such as fuel and coal combustion, vehicular and industrial emission [91]. Cd was most likely from oil leakage from automobiles along with car abrasion and car lubricant [92]. Zn might produce from the serious wear and tear of tires and the brake lining [90, 93]. Many studies reported zinc present in the rainwater due to the roofing material [94-96]. This factor further exhibits the main sources of trace metals in this study was emitted from the anthropogenic activities.

Formate, methasulfonic acid, and oxalate were the significant factor loadings in the fourth factor. It was known as organic acid, and it explained 7.02% of the variance. These organic acid loaded on the same factor may explain that it was originated from marine aerosol [97]. The high loading of oxalate (0.80) in this factor might be experienced by forming oxalate via in-situ photochemical reactions and their precursor, such as fatty acid [98]. In line with past
studies, the oxalate was the most abundant water-soluble species detected in the marine atmosphere [99, 100]. As opposed to oxalate, it originated from multiple sources, and methane sulfonic acid (MSA) was only oxidized from DMS in the air [101]. DMS was a primary source of biogenic sulfur, emitted in a large amount into the marine atmosphere [102]. MSA was detected in the atmosphere over coastal and regions [103]. While the sources of formate in this study possibly mainly from the daytime of photochemical reactions [104]. The result explained the moderate positive loading of MSA couple with formate in this factor, which was not high as oxalate. This factor concludes the sources of organic acid mainly originated from the photochemical process.

The fifth factor exhibited the strong positive loading only for manganese with 5.04% of the total variance. The emission of Mn into the atmosphere was mostly a mixture of natural crustal materials and anthropogenic compounds produced by traffic and industry [105]. This study shows Mn was not significantly correlated with other metals. It was probable that Mn originated mainly from the natural sources, particularly from soil [106]. However, this study also revealed that the traffic density was also contributed to the sources of Mn [107]. The previous studies stated that the concentration of Mn in the atmosphere resulting from the emission of dust in by vehicular emission [92]. Relatively, the high loading of Mn alone in this factor is interpreted to be the result of natural enrichment by weathering and pedogenesis [108].

The sixth factor accumulated carboxylic acid (acetate and formate), accounted for around 4.98% of the variability. It was distinguished with high loading from acetate (0.88) and moderate loading on formate (0.54). Therefore, this factor is related to a wide variety of sources, including direct emissions from biomass burning, vegetation, vehicular emissions, and urban and coastal
areas [109]. The previous study reported these carboxylic acids were the most abundant monocarboxylic acids in dust particles [110]. The higher loading of acetate compared to formate in this factor could have mainly from primary sources such as industrial and vehicular emission [104], compared to formate, possibly dominated by photochemical production [110].

The seventh factor explains 4.69% of the variance consist of negative loads of iron but overloaded with pH. Simultaneously, the eighth factor accounted for 4.56% of the total variance with positive loadings on fluoride and nitrate. Considering the value of variance for factor 6 to factor 8 less than 5.0%, this study concludes that the underlying construct of rainfall data in Eastern Malaysia was associated with five, rather than eight, factors accounting for approximately 53.38% of the variations in the dataset. Previous studies supported this finding that the variance with less than 5.0% was not significant [111]. It was associated with the criteria of factor extraction [112].

Consequently, this study did not detail out the discussion for the seventh and eighth factors. Therefore, the variables of iron and pH (factor 7) and fluoride (factor 8) were not considerably significant in this study. Though nitrate was loaded in factor 8, but it still needs to be accounted for since it falls in factor two. Based on the result of PCA, most factors in this study were dominated by water-soluble ions. These ionic species are the major components of atmospheric aerosol and are associated with acidification of rainwater [113]. The presence of these mineral components can be attributed to both natural and anthropogenic processes [82].

Table 3. Factor Loadings for the Variables after Varimax Rotation

|       | PC1 | PC2 | PC3 | PC4 | PC5 | PC6 | PC7 | PC8 |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|

20
3.3. Discrimination of Spatial Variation

In this study, DA was applied to investigate the spatial variation of rainwater characteristics between the rain gauge stations. In this study, there are three spatial groups 1, 2, and 3 represented good purity rainfall index (GPRI), moderate purity rainfall index (MPRI), and bad purity rainfall index (BPRI), respectively were treated as dependent variables. In contrast, 23 physicochemical and metal parameters were treated as independent variables. DA was performed via standard, forward, and backward stepwise modes. The accuracy of the spatial classifications through three (3) modes; standard, forward, and backward stepwise gave the values of 94.25%
with 20, 15, and 16 discriminant variables. The result of forward stepwise mode gave the highest correct classification of 94.84%, discriminating three spatial classes assignations with 15 variables, magnesium, sulfate, formate, lead, zinc, nitrate, calcium, sodium, oxalate, acetate, ammonium, manganese, meth sulfonic acid, cadmium, and pH (Table 4). Considering the high value of the correct classification matrix, the DA has substantially proven supporting the result of in confirming the PRI for each rain gauge station.

The Wilk's lambda test for the forward stepwise mode gave a value of 0.10 (p<0.0001) and shows significant differences between these three groups (Fig. 1). The null hypothesis states, the mean vectors of the three classes (GPRI, MPRI, and BPRI) are equal. The alternative hypothesis states at least one of the mean vectors is different from another. Since the computed p-value is lower than the significance level alpha = 0.05, one should reject the null hypothesis and accept the alternative hypothesis. The risk of rejecting the null hypothesis when it is true is lower than 0.01%. Thus, the three groups are indeed different from one another. Fig. 3 shows that the observations on the factor axes confirm that the rain gauge stations are very well differentiated on the factor axes extracted from the original explanatory variables. The spatial variation of rainwater characteristics results showed that discriminant analysis has successfully grouped the rain gauge stations, according to their chemical compounds in the rainwater.
Table 4. Spatial Classification Matrix of DA for All Data Measurements in Rainfall Data

| Index Scale        | % Correct | Predicted Index Scale |
|--------------------|-----------|-----------------------|
|                    | Good      | Moderate | Bad      |
| Standard Mode      |           |          |          |
| Good               | 99.48     | 386      | 2        | 0        |
| Moderate           | 73.20     | 25       | 71       | 1        |
| Bad                | 73.68     | 0        | 5        | 14       |
| Total              | 93.45     | 411      | 78       | 15       |
| Forward stepwise Mode |         |          |          |
| Good               | 99.48     | 386      | 2        | 0        |
| Moderate           | 77.32     | 21       | 75       | 1        |
| Bad                | 89.47     | 0        | 2        | 17       |
| Total              | 94.84     | 407      | 79       | 18       |
| Backward stepwise Mode |       |          |          |
| Good               | 99.48     | 386      | 2        | 0        |
| Moderate           | 75.26     | 23       | 73       | 1        |
| Good               | 89.47     | 0        | 2        | 17       |
| Total              | 94.44     | 409      | 77       | 18       |

Fig. 3. DA plot for BPRI, MRI, and BPRI.
3.4. Spatial Similarities and Site Grouping in Different Regions

HACA has been proved useful in solving classification problems where the object is the sort of variables into groups such that the degree association is strong among the members in the same class and weak among the members of different classes. In this study, HACA was performed on 23 physicochemical and metal parameters to study the spatial variations of the rain gauge stations based on their similarity levels. The result was presented in the dendrogram, as shown in Fig. 4. Six cities covering 504 stations in Labuan, Danum Valley, Kuching, Tawau, Kota Kinabalu, and Bintulu were diversified into three clusters (Fig. 2). Cluster 1 was formed Labuan and Danum Valley, named as good purity of rainfall (GPRI). Cluster 2 accommodates the urban areas namely, Kuching and Tawau as moderate purity of rainfall (MPRI). While cluster 3 formed bad purity of rainfall (BPRI) for the areas of Kota Kinabalu and Bintulu.

![Dendrogram](image)

**Fig. 4.** The dendrogram for clustering the rain gauge stations.
3.4.1. Good purity of rainfall index (Labuan and Danum Valley)

Cluster 1 corresponds to category of good purity rainfall index (GPRI) and is characterized by the biggest Euclidean distance than the other two clusters. Labuan and Danum Valley fall under GPRI. The location of rain gauge stations in these districts, mostly located in rural areas, covered with forest reserves, or protected areas such as Danum Valley. Danum Valley, situated downstream of Sabah, was far from the major point and non-point source pollution in the Sabah region. Danum valley is a forest conservation area in Sabah that covers approximately 438 km² of protected unlogged forests and is surrounded by an extensive area of production forest [114]. It is the largest remaining stand of lowland dipterocarp rain forest in Sabah with limited human settlement, least human impacts on soil and water [115]. On the site, Labuan Island, is a federal territory of Sabah, (off the coast of the state of Sabah) was located at the northwestern tip of Borneo Island [116]. Although known as being offshore support for the deep-water oil and gas industry, some small industrial area, and tourism activity [117], the result showed that Labuan had similar rainwater characteristics with Danum Valley. The minimal impact from the urbanization area, which affected by industrialization activities not contributing much to the rainwater behavior in these two districts.

3.4.2. Moderate purity of rainfall (Kuching and Tawau)

Cluster 2 corresponds to MPRI (including Kuching and Tawau) shows both regions located nearby to the coastal, industrial, and residential areas subject to the similarity amongst the rain gauge stations. Kuching is the capital of the state of Sarawak, located in the northwestern part of
Borneo. Meanwhile, Tawau is located on the Southeast coast of northern Borneo, which overlooking the Celebes Sea to the east [118]. There was substantial of low pH level detected at these districts, because of close proximity to the coastal area, with the relatively great influence of marine aerosols on the particle concentration [119]. The result has strengthened the findings of the dominance of sea salt and organic acid species in the atmosphere. Conversely, the concentration of trace metals found in this cluster to be moderately high contributed from the anthropogenic activities and natural process (river discharge, hydrothermal-circulation) as Kuching and Tawau are the major urban centre in Sarawak and Sabah, respectively [120]. The industrial areas such as Kuching City responsible for the emission of trace metals at large. The combustion of fuels from the automobiles and aircraft has become another factor of heavy metals emission in this cluster. The power plant stations located in Kuching and Tawau are the major supplier of electricity used diesel and coal-fired are the important sources of distribution and emission of hazardous trace including metal in the atmospheric environment [121]. Moreover, the precursor gases of secondary aerosols found to be predominantly originated from burning fossil fuels and motor vehicles [122]. Kuching and Tawau, both cities have good infrastructures, and domestic river port, of which enhanced the congested land and water transportation leading to the vehicle emission particles to the atmosphere [123].

3.4.3. Bad purity of rainfall (Kota Kinabalu and Bintulu)

Cluster 3 corresponds under category bad purity index (BPRI) that consists of Kota Kinabalu (Sabah) and Bintulu (Sarawak). The rain gauge stations in cluster 3 are placed in the city centers of Kota Kinabalu and Bintulu. These stations are known as major urban areas, surrounded by a
residential area, heavy traffic congestion around the clock, airport, and industrial activities [117]. Kota Kinabalu is the capital of Sabah state, located in the north of Borneo, and it is the busiest urban center in Sabah [118]. The abundance of sea salt particles ($\text{Mg}^+$, Na, and $\text{Cl}^-$) and organic acids (formate, methane sulfonic acid, and oxalate) recorded at both districts are expected due to geographical position, located near to the coastal area of which confronting the South China Sea. Lewandowska & Falkowska [124] agreed that the evaporation of sea surface sprays emitted a large amount of sea salt to the atmosphere.

The major source of heavy metals within these two areas was primarily from coal combustion and industrial emission in the region [91]. Since Kota Kinabalu and Bintulu have several stationary sources linked to the industrial such as power stations, industrial fuel combustion, and domestic fuel combustion, therefore it can be concluded that the large emissions of the precursors of inorganic ions (ammonium, nitrate, and sulfate), have been released from the industrial activities. Besides, the port activities like shipping transportation, the airport also contributes to the emission of these precursor gases. The fuel oil power plant located in this industrial area was also an essential contributor to Ni particulate levels observed at the rain gauge stations. Both Kota Kinabalu and Bintulu have some power plant stations to generate electricity in Sabah and Sarawak. Trace metals (Cu, Cd, Pb, and Zn) are highly correlated in factor 3 (trace metal), usually associated with different sources nominally coal combustion, incineration, and traffic (mainly tires and brake wear). The busiest traffic road congested with automobiles and shipping in Kota Kinabalu and Bintulu contributed to releasing trace metals into the surrounding atmosphere. The combustion of fuels from vehicular, aircraft, and the burning of fuels in industrial activities was also suspected to be the source of heavy metals. A study by
Johansson et al. [125] reported that emissions from road traffic occur at ground level in the most densely populated part of the districts were the major sources of heavy metals (Cu, Cd, Pb, Ni, Mn, and Zn). This study, in tandem with Furusjö et al. [126] and Wahlin et al. [127], observed that brake wear was the major source of heavy metal in Sweden and Copenhagen, respectively. Abbasi & Abbasi [128] agreed that the concentration of heavy metals is highest in the urban and industrial areas. Besides, the location is in the coastal zone, faced the pollution affected due to urbanization, development, and an increase in population resulting in the worst deterioration of rainwater quality than the other districts in the study area.

From the HACA result, all districts located nearby to the coastal area except Danum Valley, it can be concluded that most of the rainwater in the study area was mostly impacted by ionic particles distributed within the marine atmospheres. Therefore, these ions did not be the primary cause of the variation of rainwater characteristics in the study area. However, the rain gauge stations in the urbanized area (surrounded by the industrial zone) received abundant heavy metal particles and precursor gases from various sources. This scenario has changed the variations of rainwater behaviour tremendously. The results indicate that HACA is useful in offering reliable classification of rainwater in the study area. Additionally, this study confirms that it is conceivable to plan an optimal monitoring strategy with fewer rain gauge stations and cost-effective [129].

4. Concluding Remarks

In East Malaysia, rainwater acidity was associated with the Mg\(^+\), Na\(^+\), and Cl\(^-\) signify that these ions mostly originated from marine aerosols. However, SO\(_4^{2-}\), NO\(_3^-\) and NH\(_4^+\) are the secondary
aerosols formed via the atmosphere’s chemical reaction and mainly originated from anthropogenic activities. The alkaline ions such as Mg\(^{2+}\), Na\(^{+}\), NH\(_4\)\(^{+}\) suggested were neutralised the acidity of rainwater. Factor analysis shows that the anthropogenic activities contributing to the emission of toxic metals (8.68%) did not significantly impact the rainwater quality. The anion and cation balance varied among the rain gauge stations based on the local activities. The result shows that, with a total variance of 53.38%, sea salt, secondary aerosols, trace metals, crustal origin, and organic acid dominated. The establishment of the purity of rainfall index generated good, moderate, and bad index categories will be the indicator to determine the quality of harvested rainwater quality in East Malaysia. With the right classification value of 94.84%, only 15 variables, including magnesium, sulphate, formate, lead, zinc, nitrate, calcium, sodium, oxalate, acetate, ammonium, manganese, methasulfonic acid, cadmium, and pH, were used to effectively differentiate between three spatial groups (GPRI, MPRI and BPRI). This study suggests that an optimum monitoring approach with fewer rain gauge stations and cost-effective monitoring is conceivable.

Acknowledgement

This study is a part of the Special Research Grant Scheme UniSZA/2017/SRGS/18. The authors would like to express their appreciation to the Department of Chemistry Malaysia and Meteorology Department, Malaysia, for the supply of data required for the completion of this study. The author wishes to thank the East Coast Environmental Research Institute (ESERI) UniSZA, for the use of their research facilities.
Author Contributions

A.S.N.F. (Ph.D) wrote the manuscript. I.A. (Senior lecturer) revised the manuscript. J.H. (Professor) assisted the statistical analysis. A.R.B. (Professor) revised the manuscript. L.F. (Senior lecturer) revised the manuscript. H.N.M. (Master's student) editing the manuscript. A.N. (Master's student) editing the manuscript. Z.M.A. (Chemist) assisted the statistical analysis. M.T.A.T. (Lecturer) editing the manuscript. H.M.H.F. (Lecturer) editing the manuscript. M.R.I.S.R. (Master's student) editing the manuscript. J.J.R.A. (Director) conducted final editing the manuscript. D.S.M. (Director) conducted final editing the manuscript.

References

[1] Meng Y, Zhao Y, Li R, et al. Characterization of inorganic ions in rainwater in the megacity of Shanghai: spatiotemporal variations and source apportionment. Atmos. Res. 2019;222:12–24.

[2] Chatterjee J, Singh SK. 87Sr/ 86Sr and major ion composition of rainwater of Ahmedabad, India: Sources of base cations. Atmos. Environ. 2012;63:60–67.

[3] Al-Khashman OA. Chemical characteristics of rainwater collected at a western site of Jordan. Atmos. Res. 2009;91(1);53–61.

[4] Rao PSP, Tiwari S, Matwale JL, et al. Sources of chemical species in rainwater during monsoon and non-monsoonal periods over two mega cities in India and dominant source region of secondary aerosols. Atmos. Environ. 2016;146:90–99.

[5] Niu H, He Y, Lu XX, et al. Chemical composition of rainwater in the Yulong Snow Mountain region, Southwestern China. Atmos. Res. 2014;144:195–206.
[6] Lu X, Li LY, Li N, Yang G, Luo D, Chen J. Chemical characteristics of spring rainwater of Xi'an city, NW China. *Atmos. Environ.* 2011;45(28):5058–5063.

[7] Despins C, Farahbakhsh K, Leidl C. Assessment of rainwater quality from rainwater harvesting systems in Ontario, Canada. *J. Water Supply Res. T.* 2009;58(2):117–134.

[8] Chubaka CE, Whiley H, Edwards JW, Ross KE. A Review of Roof Harvested Rainwater in Australia. *J. Environ. Public Health.* 2018;2018:1-14.

[9] Mouli PC, Mohan SV, Reddy SJ. Rainwater chemistry at a regional representative urban site : influence of terrestrial sources on ionic composition. *Atmos. Environ.* 2005;39:999–1008.

[10] Ayob S, Rahmat SN. Rainwater harvesting (RWH) and groundwater potential as alternatives water resources in Malaysia: a review. *MATEC Web Conf.* 2017;103:04020.

[11] Handia L, Tembo JM, Mwiindwa C. Potential of rainwater harvesting in urban Zambia. *Phys. Chem. Earth.* 2003;28(20–27):893–896.

[12] Abdul Ghani NAA, Mohamad NA, Hui TW. Rainfall analysis to determine the potential of rainwater harvesting site in Kuantan, Pahang. *ARPN J. Eng. Appl. Sci.* 2016;11(11):7264–7268.

[13] Fewkes A. A review of rainwater harvesting in the UK. *Struct. Surv.* 2012;30(2):174–194.

[14] Yousif M, Bubenzer O. Geoinformatics application for assessing the potential of rainwater harvesting in arid regions. Case study: El Daba'a area, Northwestern Coast of Egypt. *Arab. J. Geosci.* 2015;8(11):9169–9191.

[15] Lani NHM, Yusop Z, Syafiuddin A. A review of rainwater harvesting in Malaysia: Prospects and challenges. *Water* 2018;10(4):1–21.

[16] Hamilton K, Reyneke B, Waso M, et al. A global review of the microbiological quality and potential health risks associated with roof-harvested rainwater tanks. *Npj Clean Water.* 2019;2(1).
[17] Reyneke B, Cloete TE, Khan S, Khan W. Rainwater harvesting solar pasteurisation treatment systems for the provision of an alternative water source in peri-urban informal settlements. *Environ. Sci. Water Res. Technol.* 2018;4:291-302.

[18] De Kwaadsteniet M, Dobrowsky PH, Van Deventer A, Khan W, Cloete TE. Domestic rainwater harvesting: Microbial and chemical water quality and point-of-use treatment systems. *Water Air Soil Pollut.* 2013;224(7):1629.

[19] Adugna D, Jensen MB, Lemma B, Gebrie GS. Assessing the potential for rooftop rainwater harvesting from large public institutions. *Int. J. Environ. Res. Public Health.* 2018;2018:15(2).

[20] Helmreich B, Horn H. Opportunities in rainwater harvesting. *Desalination* 2009;248:118–124.

[21] Zhu K, Zhang L, Hart W, Liu M, Chen H. Quality issues in harvested rainwater in arid and semi-arid Loess Plateau of northern China. *J. Arid Environ.* 2014;57(4):487–505.

[22] Villarreal EL, Dixon, A. Analysis of a rainwater collection system for domestic water supply in Ringdansen, Norrköping, Sweden. *Build. Environ.* 2005;40(9):1174–1184.

[23] Mohammed TA, Johari MMNM, Ghazali Abd Halim. Study on Potential Uses of Rainwater Harvesting in Urban areas. *Proc. Colloq. Rainwater Util.* 2007.

[24] Liaw CH., Tsai YL. Optimum storage volume of rooftop rain water harvesting systems for domestic use. *J. Am. Water Resour. Assoc.* 2004;40(4):901–912.

[25] Kim R.-H, Lee S, Kim J-O. Application of a metal membrane for rainwater utilisation: filtration characteristics and membrane fouling. *Desalination* 2005;177:121–132.
[26] Sturm M, Zimmermann M, Schütz K, Urban W, Hartung H. Rainwater harvesting as an alternative water resource in rural sites in central northern Namibia. *Phys. Chem. Earth.* 2009;34(13–16):776–785.

[27] Mansoor Ahammed M, Meera V. Iron hydroxide-coated sand filter for household drinking water from roof-harvested rainwater. *J. Water Supply Res. T.* 2006;55:493–498.

[28] Vargas D, Dominguez I, Ward S. Assisting global rainwater harvesting practitioners: selection under uncertainty. *Environ. Sci. Water Res. Technol.* 2019;3:506-520.

[29] Leeves G, Soyiri I. Does more education always lead to better health? evidence from rural Malaysia. *Hindawi* 2015;2015:539212.

[30] Yusri MSNC, Othman Z. Impact of land development on water quantity and water quality in Peninsular Malaysia. *Malaysian J. Environ. Manag.* 2011;12(2):113–120.

[31] Md Hashim N, Man S. Social perception, awareness and motivation of rainwater utilisation as an alternative supply in urban area. *Malays. J. Soc. Space.* 2018;14(3):37–52.

[32] Ghisi E, Bressan DL, Martini M. Rainwater tank capacity and potential for potable water savings by using rainwater in the residential sector of southeastern Brazil. *Int. J. Sustain. Built. Environ.* 2007;42(4):1654–1666.

[33] Rasyikah MK. Review of the water supply management and reforms needed to ensure watersecurity in Malaysia. *Int. J. Bus. Soc.* 2018;19(S3):472–483.

[34] Strauss A, Reyneke B, Waso M, Khan W. Environmental Science system for the treatment of harvested rainwater. *Environ. Sci. Water Res. Technol.* 2018;4(7):976–991.
[35] Lee KE, Mokhtar M, Mohd Hanafiah M, Abdul Halim A, Badusah J. Rainwater harvesting as an alternative water resource in Malaysia: Potential, policies and development. *J. Clean. Prod.* 2016;126:218–222.

[36] Yu J, Yan C, Liu Y, Li X, Zhou T, Zheng M. Potassium: A tracer for biomass burning in Beijing? *Aerosol. Air Qual. Res.* 2018;18(9):2447–2459.

[37] Demirel Z, Guler C. Hydrogeochemical evolution of groundwater in a Mediterranean coastal aquifer, Mersin-Erdemli basin (Turkey). *Environ. Geol.* 2006;49:477–487.

[38] Hashim H, Hudzori A, Yusop Z, Ho WS. Simulation based programming for optimisation of large-scale rainwater harvesting system: Malaysia case study. *Resour. Conserv. Recycl.* 2013;80(1):1–9.

[39] Asman NS, Halim AA, Hanafiah MM, Ariffin FD. Penentuan kualiti air daripada sistem penuaian air hujan di Kolej Ungku Omar, UKM Bangi. *Sains Malays.* 2017;46(8):1211–1219.

[40] Che-Ani AI, Shaari N, Sairi A, Zain MFM, Tahir MM. Rainwater harvesting as an alternative water supply in the future. *Eur. J. Sci. Res.* 2009;34(1):132–140.

[41] Khoon SH, Issabayeva GI, Lee LW, Sites AS. Measurement of Rainwater Chemical Composition in Malaysia based on Ion Chromatography Method. *World Acad. Sci. Eng. Technol.* 2011;5(11):2120–2127.

[42] Sumari SM, Darus FM, Kantasamy N, Taib NI, Sinyaw U, Othman IR. Compositions of Rainwater and Aerosols at Global Atmospheric Watch in Danum Valley, Sabah. *Malaysian J. Anal. Sc.* 2009;13(1):107–119.

[43] Mohammed TA, Al-molhem YA, Noor MJMM. Rainwater harvesting: quantitative and qualitative analysis. *Int. J. Eng. Technol.* 2013;10(2):62–66.
1 [44] Rasyikah MK. Review of the water supply management and reforms needed to ensure watersecurity in Malaysia. *Int. J. Bus. Soc*. 2018;19(S3):472–483.

2 [45] Juahir H, Ghazali A, Ismail, A. et al. The assessment of Danau Kota Lake water quality using chemometrics approach. *IOP Conf. Ser. Mater. Sci. Eng*. 2019;621:012019.

3 [46] Ismail A, Toriman MEM, Juahir H. et al. Chemometric techniques in oil classification from oil spill fingerprinting. *Mar. Pollut. Bull*. 2016;111(1–2):339–346.

4 [47] Ogwueleka TC. Use of multivariate statistical techniques for the evaluation of temporal and spatial variations in water quality of the Kaduna River, Nigeria. *Environ. Monit. Assess*. 2015;187(3).

5 [48] Monjerezi M, Vogt RD, Aagaard P, Saka JDK. Hydro-geochemical processes in an area with saline groundwater in lower Shire River valley, Malawi: an integrated application of hierarchical cluster and principal component analyses. *J. Appl. Geochem*. 2011;26:1399–1413.

6 [49] Sabrina IA, Sa'ezan NS. The application of rainwater as architectural design elements for green technology solution in low rise office building. *Int. J. Sustain*. 2018;5(3):214–223.

7 [50] Syafrina AH, Zalina MD, Juneng L. Historical trend of hourly extreme rainfall in Peninsular Malaysia. *Theor. Appl. Climatol*. 2015;120(1–2):259–285.

8 [51] Semire FA, Mohd MR., Ismail W, Mohamad N, Mandeep JS. Ground validation of space-borne satellite rainfall products in Malaysia. *Adv. Space Res*. 2012;50(9):1241–1249.

9 [52] Inostroza L, Tábbita JH. Informal urban development in the greater buenos aires area: a quantitative-spatial assessment based on households' physical features using gis and principal component analysis. *Procedia Eng*. 2016;161(0):2138-2146.
[53] Singh KP, Malik A, Mohan D, Sinha S. Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India) - a case study. *Water Res.* 2004;38(18):3980–3992.

[54] Cloutier V, Lefebvre R, Therrien R, Savard MM. Multivariate statistical analysis of geochemical data as indicative of the hydrogeochemical evolution of groundwater in a sedimentary rock aquifer system. *J. Hydrol.* 2008;353:294–313.

[55] Liu C, Lin K, Kuo Y. Application of factor analysis in the assessment of groundwater quality in a blackfoot disease area in Taiwan. *Sci. Total Environ.* 2003;313:77–89.

[56] Menció A, Mas-Pla J. Assessment by multivariate analysis of groundwater-surface water interactions in urbanised Mediterranean streams. *J. Hydrol.* 2008;352:355–366.

[57] Fazillah A, Toriman E, Juahir H, Izwan A, Mohamad N, Ismail A. Evaluation of socioeconomic status on drug addicts. *Int. J. Adv. Sci. Eng. Inf. Technol.* 2017;7(6):2122–2130.

[58] Zolkipli HM, Juahir H, Adiana G, Zainuddin N, Ismail A, Maliki ABHM, Hussain NI, Kamarudin M KA, Toriman ME, Mokhtar M. Spatial assessment and the most significant parameters for drinking water quality using chemometric technique: A case study at Malaysia water treatment plants. *Int. J. Eng. Technol.* 2018;7(14):115–122.

[59] Kannel PR, Lee S, Kanel SR, Khan SP. Chemometric application in classification and assessment of monitoring locations of an urban river system. *Anal. Chim. Acta.* 2017;582:390–399.

[60] Wu M-L, Wang YS, Sun CC. et al. Identification of coastal water quality by statistical analysis methods in Daya Bay, South China Sea. *Mar. Pollut. Bull.* 2010;60(6):852–860.
[61] Toriman ME, Abdullah SNF, Azizan IA, Kamarudin MKA, Umar R, Mohamad N. Spatial and temporal assessment on drug addiction using multivariate analysis and GIS. *Malaysian J. Anal. Sci.* 2015;19(6):1361–1373.

[62] Khairuddin Z, Juahir H, Endut A, Ismail A, Kamarudin MKA, Ghazali A, Mokhtar M, Toriman M. E, Islam MS, Lananan F, Hassan AR. Moringa Oleifera dosage clustering for remediation process of Batik effluents using chemometric technique. *Int. J. Eng. Technol.* 2018;7(14):85–89.

[63] Alberto WD, María del Pilar D, Maria Valeria A, Fabiana PS, Cecilia HA, María de los Ángeles B. Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality. a case study: Suquía River Basin (Córdoba–Argentina). *Water Res.* 2001;35(12):2881–2894.

[64] Liu M, Huang X, Song Y. et al. Ammonia emission control in China would mitigate haze pollution and nitrogen deposition, but worsen acid rain. *Proc. Natl. Acad. Sci. U S A.* 2019;116(16):7760–7765.

[65] Zhang L, Chen Y, Zhao Y, et al. Agricultural ammonia emissions in China: reconciling bottom-up and top-down estimates. *Atmos. Chem. Phys.* 2018;339–355.

[66] Raman RS, Hopke PK, Holsen TM. Characterisation of fine aerosol and its inorganic components at two rural locations in New York State. *Environ. Monit. Assess.* 2008;144(1–3):351–366.

[67] Wen W, Cheng S, Liu L. et al. PM2.5 chemical composition analysis in different functional subdivisions in Tangshan, China. *Aerosol. Air Qual. Res.* 2016;16(7):1651–1664.
[68] Chen J, Li F, Fan Z, Wang Y. Integrated application of multivariate statistical methods to source apportionment of watercourses in the liao river basin, Northeast China. *Int. J. Environ. Res. Public Health.* 2016;13(10):1035–1061.

[69] Santos M, Dos Dawidowski L, Smichowski P, Ulke AG, Gómez D. Factors controlling sea salt abundances in the urban atmosphere of a coastal South American megacity. *Atmos. Environ.* 2012;59:483–491. https://doi.org/10.1016/j.atmosenv.2012.05.019.

[70] Cempel M, Nikel G. Nickel: a review of its sources and environmental toxicology. *Pol. J. Environ. Stud.* 2006;15(3):375–382.

[71] Tian HZ, Lu L, Cheng K. et al. Anthropogenic atmospheric nickel emissions and its distribution characteristics in China. *Sci. Total Environ.* 2012;417–418:148–157.

[72] Xu H, Ho SSH, Cao J. et al. A 10-year observation of PM 2.5-bound nickel in Xi'an, China: Effects of source control on its trend and associated health risks. *Sci. Rep.* 2017;71–10.

[73] Harasim P, Filipek T. Nickel in the environment. *J. Elem.* 2015;20(2):525–534.

[74] Zhang T, Cao JJ, Tie XX, et al. Water-soluble ions in atmospheric aerosols measured in Xi'an, China: seasonal variations and sources. *Atmos. Res.* 2011;102(1–2):110–119.

[75] Yao X, Lau APS, Fang M, Chan CK, Hu M. Size distributions and formation of ionic species in atmospheric particulate pollutants in Beijing, China: 1 — inorganic ions. *Atmos. Environ.* 2003;37:2991–3000.

[76] Ianniello A, Spataro F, Esposito G, Allegrini I, Hu M, Zhu T. Chemical characteristics of inorganic ammonium salts in PM2.5 in the atmosphere of Beijing (China). *Atmos. Chem. Phys.* 2011;11(21):10803–10822.
[77] Zhang H, Li J, Ying Q. et al. Source apportionment of PM 2.5 nitrate and sulfate in China using a source-oriented chemical transport model. *Atmos. Environ.* 2012;62:228–242.

[78] Bencs L, Ravindra K, De Hoog J. et al. Mass and ionic composition of atmospheric fine particles over Belgium and their relation with gaseous air pollutants. *Environ. Monit.* 2008;10(10):1148–1157.

[79] Allen AG, Cardoso AA, Da Rocha GO. Influence of sugar cane burning on aerosol soluble ion composition in Southeastern Brazil. *Atmos. Environ.* 2004;38(30):5025–5038.

[80] Farren NJ, Dunmore RE, Mead, M. I. et al. Chemical characterisation of water-soluble ions in atmospheric particulate matter on the east coast of Peninsular Malaysia. *Atmos. Chem. Phys.* 2019;19(3):1537–1553.

[81] Chan CK, Yao X. Air pollution in mega cities in China. *Atmos. Environ.* 2008;42:1–42.

[82] Miranda R, Tomaz E. Characterization of urban aerosol in Campinas, São Paulo, Brazil. *Atmos. Res.* 2008;87(2):147–157.

[83] Sequeira R, Lai K. The effect of meteorological parameters and aerosols constituents on visibility in urban Hong Kong. *Atmos. Environ.* 1998;32(16):2865–2871.

[84] Bourotte C, Curi-amarante A, Forti M. Association between ionic composition of fine and coarse aerosol soluble fraction and peak expiratory flow of asthmatic patients in São Paulo city (Brazil). *Atmos. Environ.* 2007;41:2036–2048.

[85] Kulshrestha UC, Kulshrestha MJ, Sekar R, Sastry GSR, Vairamani M. Chemical characteristics of rainwater at an urban site of south-central India. *Atmos. Environ.* 2003;37(21):3019–3026.
[86] Chakraborty A, Gupta T. Chemical characterisation and source apportionment of submicron (PM 1) aerosol in Kanpur Region, India. *Aerosol Air Qual. Res.* 2010;10(5):433–445.

[87] Wang Y, Zhuang G, Zhang X. et al. The ion chemistry, seasonal cycle, and sources of PM 2.5 and TSP aerosol in Shanghai. *Atmos. Environ.* 2006;40(16):2935–2952.

[88] Kabadayi F, Cesur H. Determination of Cu, Pb, Zn, Ni, Co, Cd, and Mn in road dusts of Samsun City. *Environ. Monit. Assess.* 2010;168(1–4):241–253.

[89] Zhang C. Using multivariate analyses and GIS to identify pollutants and their spatial patterns in urban soils in Galway, Ireland. *Environ. Pollut.* 2006;142(3):501–511.

[90] Zhong JNM, Latif MT, Mohamad N, Wahid NBA, Dominick D, Juahir H. Source apportionment of particulate matter (pm 10) and indoor dust in a university building. *Environ. Forensics.* 2014;15(1):8–16.

[91] Suvarapu LN, Baek SO. Determination of heavy metals in the ambient atmosphere: A review. *Toxicol. Ind. Health.* 2016;33(1):79–96. https://doi.org/10.1177/0748233716654827.

[92] Munim N, Talib M, Othman M, Dominick D, Mohamad N, Juahir H, Tahir NM. Composition of selected heavy metals in road dust from Kuala Lumpur city centre. *Environ. Earth Sci.* 2014;(72):849-859.

[93] Apeagyei E, Bank MS, Spengler JD. Distribution of heavy metals in road dust along an urban-rural gradient in Massachusetts. *Atmos. Environ.* 2011;45(13):2310–2323.

[94] De Kwaadsteniet M, Dobrowsky PH, Van Deventer A, Khan W, Cloete TE. Domestic rainwater harvesting: Microbial and chemical water quality and point-of-use treatment systems. *Water Air Soil Pollut.* 2013;224(7).
[95] Huston R, Chan YC, Chapman H, Gardner T, Shaw G. Source apportionment of heavy metals and ionic contaminants in rainwater tanks in a subtropical urban area in Australia. *Water Res.* 2012;46(4):1121–1132.

[96] Lee JY, Kim HJ, Han MY. Quality assessment of rooftop runoff and harvested rainwater from a building catchment. *Water Sci. Technol.* 2011;63(11):2725–2731.

[97] Miyazaki Y, Kawamura K, Sawano M. Size distributions and chemical characterisation of water-soluble organic aerosols over the western North Pacific in summer. *J. Geophys. Res. Atmos.* 2010;115(23):1–13.

[98] King CF, Thomas ER, Pedro JB. et al. Organic compounds in a sub-Antarctic ice core: a potential suite of sea ice markers. *Geophys. Res. Lett.* 2019;9930–9939.

[99] Wang H, Kawamura K, Yamazaki K. Water-soluble dicarboxylic acids, ketoacids and dicarbonyls in the atmospheric aerosols over the Southern Ocean and Western Pacific Ocean. *J. Atmos. Chem.* 2006;53(1):43–61.

[100] Kawamura K, Umemoto N, Mochida M, Bertram T, Howell S, Huebert BJ. Water-soluble dicarboxylic acids in the tropospheric aerosols collected over east Asia and western North Pacific by ACE-Asia C-130 aircraft. *J. Geophys. Res. Atmos.* 2003;108(23):1–7.

[101] Fuooki A, Miura K, Uematsu M. The increase of biogenic sulfate aerosol and particle number in marine atmosphere over the Northwestern North Pacific. *J. Oceanogr.* 2003;59:799–807.

[102] Fu P, Kawamura K, Chen J. et al. Fluorescent water-soluble organic aerosols in the High Arctic atmosphere. *Sci. Rep.* 2015;5:9845.
[103] Kwong KC, Chim MM, Hoffmann EH. et al. Chemical transformation of methanesulfonic acid and sodium methanesulfonate through heterogeneous OH oxidation. *ACS Earth Space Chem.* 2018;2:895-903..

[104] Mkoma SL, da Rocha GO, Domingos JSS. et al. Atmospheric particle dry deposition of major ions to the South Atlantic coastal area observed at Baía de Todos os Santos, Brazil. *An. Acad. Bras. Ciênc.* 2014;86(1):37–55.

[105] Lu X, Wang L, Lei K, Huang J, Zhai Y. Contamination assessment of copper, lead, zinc, manganese and nickel in street dust of Baoji, NW China. *J. Hazard Mater.* 2009;161:1058–1062.

[106] Yongming H, Peixuan D, Junji C, Posmentier ES. Multivariate analysis of heavy metal contamination in urban dusts of Xi’an, Central China. *Sci. Total Environ.* 2006;355:176–186.

[107] Moreno T, Pandolfi M, Querol X, Lavín J, Alastuey A, Viana M, Gibbons W. Manganese in the urban atmosphere: identifying anomalous concentrations and sources. *Environ. Sci. Pollut. Res.* 2011;18(2):173–183.

[108] Al-Khashman OA. The investigation of metal concentrations in street dust samples in Aqaba city, Jordan. *Environ. Geochem. Health.* 2007;29(3):197–207.

[109] Boreddy SKR, Mochizuki T, Kawamura K, Bikkina S, Sarin MM. Homologous series of low molecular weight (C1-C10) monocarboxylic acids, benzoic acid and hydroxyacids in fine-mode (PM2.5) aerosols over the Bay of Bengal: influence of heterogeneity in air masses and formation pathways. *Atmos. Environ.* 2017;167:170–180.

[110] Paulot F, Wunch D, Crounse JD. et al. Importance of secondary sources in the atmospheric budgets of formic and acetic acids. *Atmos. Chem. Phys.* 2011;11(5):1989–2013.
[111] Gazzaz NM, Yusoff MK, Ramli MF, Aris AZ, Juahir H. Characterization of spatial patterns in river water quality using chemometric pattern recognition techniques. *Mar. Pollut. Bull.* 2012;64(4):688–698.

[112] Reimann C, Filzmoser P, Garrett RG. Factor analysis applied to regional geochemical data: Problems and possibilities. *J. Appl. Geochem.* 2002;17(3):185–206.

[113] Nicolás JF, Galindo N, Yubero E, Pastor C, Crespo J, Esclapez R. Aerosol inorganic ions in a semiarid region on the southeastern Spanish Mediterranean coast. *Water Air Soil Pollut.* 2009;201:149–159.

[114] Pyle JA, Warwick NJ, Harris NRP. et al. The impact of local surface changes in Borneo on atmospheric composition at wider spatial scales: coastal processes, land-use change and air quality. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 2011;366(1582):3210–3224.

[115] Colón CP. Ranging behaviour and activity of the Malay civet (Viverra tangalunga) in a logged and an unlogged forest in Danum Valley, East Malaysia. *J. Zool.* 2002;257(4):473–485.

[116] Ponnan R. Broadcasting and socially responsible rural tourism in Labuan, Malaysia. *Worldw. Hosp. Tour. Themes.* 2013;5(4):398–411.

[117] Chang HWJ, Chee FP, Kong SK, Sentian J. Variability of the PM10 concentration in the urban atmosphere of Sabah and its responses to diurnal and weekly changes of CO, NO2, SO2 and Ozone. *Asian J. Atmos. Environ.* 2018;12(2):109–126.

[118] Dominick D, Juahir H, Latif MT, Zain SM, Aris AZ. Spatial assessment of air quality patterns in Malaysia using multivariate analysis. *Atmos. Environ.* 2012;60:172–181.
[119] Almeida SM, Pio CA, Freitas MC, Rei, MA, Trancoso MA. Source apportionment of fine and coarse particulate matter in a sub-urban area at the Western European Coast. *Atmos. Environ.* 2005;39:3127–3138.

[120] Hasan AR, Nair PL. Urbanisation and growth of metropolitan centres in Malaysia. *Malays. J. Econ. Studies.* 2014;51(1):87–101.

[121] Ahmad NA, Abdul-Ghani AA. Towards sustainable development in Malaysia: In the perspective of energy security for buildings. *Procedia Eng.* 2011;20:222–229.

[122] Mittal ML, Hess PG, Jain SL, Arya BC, Sharma C. Surface ozone in the Indian region. *Atmos. Environ.* 2007;41(x):6572–6584.

[123] Norsukhairin S, Abdul S, Azid A, et al. Spatial and temporal air quality pattern recognition using environmetric techniques: a case study in Malaysia. *Environ. Sci-Proc. Imp.* 2013;15:1717–1728.

[124] Lewandowska AU, Falkowska LM. Sea salt in aerosols over the southern Baltic. Part 1. The generation and transportation of marine particles. *Oceanologia.* 2013;55(2):279–298.

[125] Johansson C, Norman M, Burman L. Road traffic emission factors for heavy metals. *Atmos. Environ.* 2009;43(31):4681–4688.

[126] Furusjö E, Sternbeck J, Cousins AP. PM 10 source characterisation at urban and highway roadside locations. *Sci. Total Environ.* 2007;387:206–219.

[127] Wahlin P, Berkowicz R, Palmgren F. Characterisation of traffic-generated particulate matter in Copenhagen. *Atmos. Environ.* 2006;40:2151–2159.

[128] Abbasi, T., & Abbasi, S. A. Sources of pollution in rooftop rainwater harvesting systems and their control. *Crit. Rev. Environ. Sci. Technol.* 2011;41(23),2097–2167.
[129] Wang Y, Wang P, Bai, Y. Assessment of surface water quality via multivariate statistical techniques: a case study of the Songhua River Harbin region, China. *J. Hydro-environ. Res.* 2013;7:30–40.