


## Article

# Characteristics of Photochemical Reactions with VOCs Using Multivariate Statistical Techniques on Data from Photochemical Assessment Monitoring Stations

Edward Ming-Yang Wu<sup>1</sup> and Shu-Lung Kuo<sup>2,\*</sup> <sup>1</sup> College of Maritime, National Kaohsiung University of Science and Technology, Kaohsiung 805301, Taiwan<sup>2</sup> Department of Technology Management, Open University of Kaohsiung, Kaohsiung 812008, Taiwan

\* Correspondence: singsuey@ms28.hinet.net

**Abstract:** This study assesses the concentrations of the 54 ozone precursors (all being volatile organic compounds (VOCs)) detected at the four photochemical assessment monitoring stations that are part of the air quality monitoring network in the Kaohsiung-Pingtung area in Taiwan. Factor and cluster analyses of the multivariate statistical analysis are performed to explore the interrelationship among the 10 VOCs of relatively higher concentrations selected from the 54 ozone precursors to identify significant factors affecting ozone pollution levels in the study area. Moreover, the multivariate statistical analysis can faithfully reflect why the study area has been affected by photochemical pollution. First, results of the factor analysis suggest that the factors affecting how photochemical reactions occur in the study area can be divided into the following: “pollution from mobile sources”, “pollution from stationary sources”, and “pollution from energy sources”. Among them, mobile sources have the greatest impact on photochemical pollution levels. Second, the impacts of photochemical pollution on air quality in the study area can be classified into four clusters via cluster analysis. Each cluster represents how the 10 VOCs affect air quality, with different characteristics, and how they contribute to photochemical pollution in the study area. If there are more types and samples of photochemical pollutants when performing a multivariate statistical analysis, the analysis results will be more stable. This study adopts data on VOC monitoring over a period of nearly two years, which can effectively improve the validity and reliability of the factor analysis results, while helping environmental agencies review the effectiveness of air quality management in the future and serving as reference for the effectiveness of reducing photochemical pollution in the atmosphere.

**Keywords:** multivariate statistical techniques; VOCs; photochemical assessment monitoring stations; photochemical reaction; ozone precursors



**Citation:** Wu, E.M.-Y.; Kuo, S.-L. Characteristics of Photochemical Reactions with VOCs Using Multivariate Statistical Techniques on Data from Photochemical Assessment Monitoring Stations. *Atmosphere* **2022**, *13*, 1489. <https://doi.org/10.3390/atmos13091489>

Academic Editors: Wen-Hsi Cheng, Chung-Shin Yuan and Yoshizumi Kajii

Received: 14 July 2022

Accepted: 10 September 2022

Published: 13 September 2022

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Photochemical pollution is a new type of pollutant that came about after petroleum was used as a fuel. It first occurred in the third largest city in the United States, Los Angeles, in 1944. Since the 1950s, photochemical pollution has become a recurrent problem across the globe [1,2]. When the primary pollutants of automotive exhaust, such as hydrocarbons and NO<sub>2</sub>, are released into the atmosphere, they absorb energy from the sun under strong ultraviolet irradiation. The molecules in said substances will become unstable after they absorb energy from the sun, the original chemical bonds will be broken, and new substances are formed. Such a chemical reaction is called a photochemical reaction, and its product is the toxic photochemical smog. In areas where there are a high number of transport vehicles and industrial districts where factories emitting nitrogen oxides and hydrocarbons are concentrated, people need to be aware of photochemical pollution [3–5].

VOCs refer to any carbon-containing volatile compounds [6]. They can be found in many products, such as solvent-based paints, printing inks, consumer goods, organic solvents, and petroleum products [6,7]. Vehicles and vessels also emit VOCs, ultimately

leading to air pollution and smog. VOCs play an important role in the formation of ozone in the atmosphere. Under sunlight, VOCs will react with the nitrogen oxides emitted from automobiles, power plants, and industrial activities to form ozone. It is clear that VOCs are harmful to human health. When the concentration of a VOC in a family room reaches a certain level, people will have headaches, nausea, vomiting, and limb weakness. In severe cases, they will have a seizure, fall into a coma, or have hypomnesia [8–10]. Their respiratory system, lungs, liver, kidneys, nervous system, hematopoietic system, and digestive system will be affected when they are exposed to an environment with high concentrations of VOCs. Some VOCs are carcinogenic. Long-term exposure to an environment with high concentrations of VOCs will increase the risk of developing cancer [8,10]. When a photochemical reaction occurs, VOCs in the atmosphere will hinder the growth of crops, including decreasing crop yields and increasing plants' susceptibility to disease [11]. Moreover, VOCs will cause pollution to the local environment as wind blows or rain falls, or they will enter bodies of water or soil when they move and diffuse.

Multivariate statistical analysis is a branch of statistics commonly used in the fields of management science, social science, and life science. It is used to analyze data containing multiple variables, explore data relevance, or clarify the data structure. It differs from traditional statistical approaches that emphasize parametric estimation and hypothesis testing [12]. As multivariate statistical analysis requires complicated and massive calculations, computers are used to perform said calculations. Among multivariate statistical analysis techniques, factor analysis aims to simplify a set of interrelated variables, turning them into a smaller number of meaningful dimensions or factors. In other words, a factor can be used to represent or replace some variables of similar nature among said variables. Thus, we hope to use a few main factors to cover a set of variables via factor analysis [13,14]. A cluster analysis classifies a group of individuals based on their variables or characteristics. When performing a cluster analysis, clusters of said individuals are not known beforehand, and individuals of similar characteristics are classified based on their variable data. Hence, as with factor analysis results, cluster analysis results are obtained by calculating the mathematical statistics of the provided data. Whether said results have substantial or theoretical meaning and how many clusters are to be adopted in the end must be judged by those performing the analysis [15–17].

This study performs factor and cluster analyses of the multivariate statistical analysis of 10 VOCs of relatively higher concentrations selected from the 54 ozone precursors whose concentrations at the four photochemical assessment monitoring stations in the study area have been assessed over the years. The goal is to properly classify the pollution caused by the emissions of the 10 VOCs. It is hoped that the exploration into the characteristics of the distribution of photochemical pollution caused by emissions of the 10 VOCs can immediately reflect the differences in the extent to which the photochemical assessment monitoring stations have been affected by photochemical pollution. Moreover, said exploration can serve as reference for environmental agencies when implementing air quality management in the future.

## 2. Methodology

### 2.1. Locations of Photochemical Assessment Monitoring Stations and Data Selection

This study selected four photochemical assessment monitoring stations set up by Taiwan's Environmental Protection Administration in Kaohsiung and Pingtung based on local geographical features (i.e., Qiaotou station, Xiaogang station, and Linyuan station in Kaohsiung City and Chaozhou station in Pingtung). The study area is located in the Pingtung Plain at an elevation of approximately 30–100 m above sea level. Its average yearly temperature is 24–25 °C. The average temperature of the warmest months (July ~ August) is above 30 °C, and that of the coldest months (January ~ February) is above 18 °C. The annual precipitation is 1500 mm–2000 mm [6,10]. The study area has a tropical monsoon climate. The hottest and most humid months of the year are between May and October. It is particularly clear that June to August comprise the rainy season, during which the

rainfall accounts for 82% of the annual rainfall in Kaohsiung. The coldest month in the study area is January, with an average low temperature of 19.9 °C; the hottest month is July, with an average high temperature of 29.4 °C. The least humid month is December, with an average relative humidity of 72.4%; the most humid months are between June and August, with an average relative humidity of 81.3%. Due to its geographical location, high population density (Kaohsiung City has a population of 2.77 million), and the fact that it is the most important heavy industrial area in Taiwan, the study area has a large number of motor vehicles and stationary sources (chimneys) at usual times. It is thus more likely to be affected by air pollution. Since Kaohsiung has a high population density and relatively frequent stationary and mobile sources, more photochemical assessment monitoring stations are located here. Ten VOCs of higher atmospheric concentrations among the 54 ozone precursors (all being VOCs), whose concentrations in the monitoring stations have been assessed over the years, are used as a basis for the analysis. Said 10 VOCs include acetylene (AT), isopentane (IP), n-butane (NB), 1,2,4-trimethylbenzene (TB), toluene (TU), propene (PP), 2-methylhexane (MH), ethane (EA), isopropyl benzene (IB), and n-undecane (NC). The monitoring and operating models of photochemical assessment monitoring stations in the USA were introduced by Taiwan's Environmental Protection Administration to the photochemical assessment monitoring stations in question, which are equipped with instruments and facilities. PerkinElmer's automatic VOC analysis system has been used as the basic monitoring facilities in the photochemical assessment monitoring stations. This analytical instrument automatically collects samples and makes measurements every hour. VOCs in the air will be retained in the electronic cooling concentration trap when the measurements are made. The trap will be filled with an appropriate adsorbent to obtain a fixed amount of cooled, concentrated C<sub>2</sub>-C<sub>11</sub> VOCs. Monitoring data collected between 1 July 2020 and 31 May 2022 were selected; 24 complete VOC data entries per day (24 × 60 days = 14,400 entries) are used as the basis for the model analysis. Figure 1 shows the geographical locations of the four photochemical assessment monitoring stations. The selection of monitoring data collected over a period of one and a half years takes into consideration trends of seasonal and time-series variations in air pollutants (e.g., different concentration levels in the mornings and evenings). Evidence-based software (SPSS Statistics 26, IBM, Armonk, NY, USA) was adopted for this study.

## 2.2. Multivariate Statistical Analyses—Factor Analysis

Factor analysis aims to identify basic variables (referring to the 10 VOCs in this study) to explain the relevant modes demonstrated in the set of observed variables. It is typically used for the dimensionality reduction of data. The goal is to identify a few factors and use them to explain the variance observed among the numerous manifest variables. The advantage of factor analysis is that it can explain most variations by means of fewer variables. With principal component analysis (PCA), these variables can be integrated into indexes to reveal the common factors from the original data that are relevant to one another [18].

In factor analyses, each variable, i.e.,  $x_1 \sim x_p$ , in a set of  $p$  variables is decomposed into  $q$  common factors, i.e.,  $f_1 \sim f_p$  ( $q \leq p$ ), which are linearly combined with the special factors  $\epsilon_i$ .

The model of factor analyses can be represented by:

$$\begin{aligned}
 x_1 &= \mu_1 + l_{11}f_1 + l_{12}f_2 + \dots + l_{1q}f_q + \epsilon_1 \\
 x_2 &= \mu_2 + l_{21}f_1 + l_{22}f_2 + \dots + l_{2q}f_q + \epsilon_2 \\
 &\vdots \\
 x_i &= \mu_i + l_{i1}f_1 + l_{i2}f_2 + \dots + l_{iq}f_q + \epsilon_i \\
 x_p &= \mu_p + l_{p1}f_1 + l_{p2}f_2 + \dots + l_{pq}f_q + \epsilon_p
 \end{aligned} \tag{1}$$

where  $f_1, \dots, f_q$  are common factors contained in each variable  $x_i$ ,  $\epsilon_i$  is special factor contained only in the  $i$ th variable ( $x_i$ ), and  $\ell_{ij}$  is the loading of  $i$ th factor to the  $j$ th common factor ( $f_j$ ).



**Figure 1.** Photochemical monitoring stations in the Kaohsiung-Pingtung area of Taiwan.

### 2.3. Multivariate Statistical Analyses—Cluster Analysis

The so-called cluster analysis is a method used to group similar samples among  $p$  samples or subjects in the same cluster based on their characteristics. As such, said samples or subjects can be classified into several clusters and characteristics of each cluster analyzed. The objective of a cluster analysis is to divide samples into different clusters to maximize the homogeneity of each cluster and the heterogeneity among the clusters. Each cluster thus describes, in terms of the data collected, the class to which its members belong, and this description may be abstracted through use from the particular to the general class or type. Hierarchical agglomerative clustering is the most common approach, providing intuitive similarity relationships between any one sample and the entire dataset, and it is typically illustrated by a dendrogram (tree diagram) [19,20]. The dendrogram provides a visual summary of the clustering processes, presenting a picture of the groups and their proximity, with a dramatic reduction in dimensionality of the original data. The Euclidean distance usually gives the similarity between two samples, and a distance can be represented by the difference between analytical values from the samples [20]. The Ward's method uses an

analysis of variance approach to evaluate the distances between clusters in an attempt to minimize the sum of squares of any two clusters that can be formed at each step.

### 3. Results and Discussion

#### 3.1. VOC Data Processing

The data on the VOCs were collected over the period from 1 July 2020 to 31 May 2022, and a total of 14,400 data entries were used. Before analyzing the sequence of the 10 VOCs using the multivariate model, this study standardized the source data since the data on the 10 VOCs possessed seasonal and periodic characteristics as follows:

$$Z_{ij} = \frac{X_{ij} - \bar{X}_i}{S_i}$$

where  $X_{ij}$  is the sequence of  $i$  VOCs and the measured value of the  $j$  sample,  $\bar{X}_i$  is the average of the sequence of  $i$  VOCs, and  $S_i$  is the standard deviation of the sequence of  $i$  VOCs.

#### 3.2. Selecting the Results of the Factor Analysis

This study explains the characteristics of factors using the varimax rotation in factor analysis to perform orthogonal rotation. The analysis results show three factors with eigenvalues of greater than 1, as shown in Table 1. The eigenvalue is used to indicate whether a factor in question is significant or not. If the eigenvalue is greater than 1, the factor is considered “significant”. Otherwise, the factor is “insignificant” or it is white noise that can be ignored. These three factors can be used to fully explain the main factors affecting the generation of photochemical pollution in the study area. The cumulative explained variance of the three common factors is 83.437%, and their eigenvalues are 2.244, 1.682, and 1.161, respectively. Table 2 shows that the KMO value is 0.876. According to Kaiser, this result is suitable for a factor analysis since it is greater than 0.5. Moreover, as shown in Table 2, the chi-square distribution value of Bartlett’s test of sphericity is 2411.390 (and the degree of freedom is 22). Having achieved the significance level, this value represents the existence of common factors among the relevant matrices of the population and is therefore suitable for a factor analysis.

**Table 1.** Results of the factor analysis and the variance explained.

Components	Initial Eigenvalues	% of Total Variance	Cumulative Variance %
1	2.244	43.478	43.478
2	1.682	26.087	69.565
3	1.161	13.872	83.437
4	0.730	7.765	91.202
5	0.506	3.641	94.843
6	0.443	2.538	97.381
7	0.324	1.009	98.390
8	0.216	0.624	99.014
9	0.203	0.553	99.567
10	0.118	0.433	100.000

**Table 2.** KMO and Bartlett’s test of the 10 VOCs.

<b>Kaiser–Meyer–Olkin Measure of Sampling Adequacy</b>		0.876
<b>Bartlett’s test of sphericity</b>	Chi-square distribution	2411.390
	Degree of freedom	22
	Significance	0.000

### 3.3. Determining the Number of Factors

The number of eigenvalues greater than 1, as shown in Table 1, can help to determine the number of main factors affecting the mechanism for the generation of photochemical pollution in the study area. Table 3 shows the component matrix after orthogonal rotation. The matrix after rotation can be used to explain the characteristics of the factors. The three factors can also be used to describe the relevance among the 10 VOCs detected in the air quality monitoring network and the impacts of said VOCs on the generation of photochemical pollution in the study area.

**Table 3.** Loading matrix of the three factors.

VOCs	Factors		
	1	2	3
TU	0.879	0.211	0.006
TB	0.792	−0.095	−0.058
IB	0.741	0.117	−0.084
IP	−0.567	0.822	0.273
AT	0.315	0.770	0.049
MH	0.098	0.713	0.224
EA	−0.284	0.164	0.858
PP	0.346	−0.263	0.780
NB	−0.148	0.096	0.733
NC	0.220	0.369	0.681

Note: These VOCs abbreviations are explained in Section 2.1.

### 3.4. Explanation of Factors

Table 3 shows the three main factors affecting the mechanism for the generation and characteristics of photochemical pollution in the study area. The following is a complete description of the characteristics of the formation of said three factors.

#### 3.4.1. Factor 1

As shown in Table 3, Factor 1 consists of three VOCs: TU, TB, and IB. Table 1 shows that their total variance can reach 43.478%. The impact of said VOCs on atmospheric photochemical pollution accounts for almost half of the total variance explained. It can be learned from Table 3 that TU, a constituent of Factor 1, has the highest loading of 0.879, followed by TB and IB with loadings of 0.792 and 0.741, respectively. The mechanism for the formation of said three VOCs is related to organic pollutants emitted by motor vehicles in the study area, which is attributable to the fact that the study area has a high population density. It not only has a huge number of permanent residents, but also of motor vehicles, especially in the Kaohsiung metropolitan area, where there is a high ratio of using motor vehicles for commuting or travel purposes during holidays. Mobile sources tend to affect photochemical pollution easily in one day [3]. The study area is the major heavy industrial area in Taiwan; thus, considerable concentrations of VOCs emitted from stationary sources are detected here [17]. In addition, stationary sources in Kaohsiung such as oil refineries, power plants, and petrochemical plants usually release VOCs of higher concentrations into the atmosphere during the daytime. There is also a large number of vehicles traveling to and from the Kaohsiung metropolitan area at the usual times and during holidays. Thus, the concentrations of VOCs are higher here during the daytime. On the contrary, the concentrations of pollution from various stationary and mobile sources are significantly lower at nighttime than those during of the daytime. In terms of trends of seasonal variations, the study area is in the northeast monsoon, wake low wind area in the winter. Air pollutants generated in the north are transported to the south by the monsoon, whereby photochemical pollution is more likely to occur. At this time, the wind gradually blows toward the east in the local atmospheric wind field. As the area is on the leeward side, it is subject to lower average wind speeds and lower rainfall levels, which makes it difficult

for VOCs to diffuse. Consequently, the VOCs contribute to air pollution for a longer period of time. However, the sources of TU, TB, and IB are the emissions from motor vehicles. Moreover, there is a considerable number of industrial areas in Kaohsiung, along with the Port of Kaohsiung, which is the 15th largest port in the world. There are large vehicles traveling to and from Kaohsiung at the usual times, resulting in high concentrations of VOCs in the atmosphere even at off-peak times. The study area has long been affected by a higher concentration of TU emitted by motor vehicles. In spring, autumn, and winter, TU in the atmosphere can easily accumulate due to changes in weather conditions, low rainfall, and other factors (such as mixing height, boundary layer height, suppressed convective mixing, etc.); consequently, it cannot be easily diffused [21]. Although the concentrations of TB and IB in the atmosphere are not as high as that of TU, they are the main types of mobile sources in the study area. Moreover, they are the main sources of photochemical pollution.

To sum up, the analysis results suggest that the main sources and emission behaviors of TU, TB, and IB detected in the study area are closely related to exhaust emissions from motor vehicles. Their molecular structure belongs in that of aromatic hydrocarbons. They also constitute the main factors that contribute to the generation of photochemical pollution in the study area. As such, Factor 1 can be called “the factor of pollution from mobile sources”.

#### 3.4.2. Factor 2

As shown in Table 3, Factor 2 consists of three VOCs: IP, AT, and MH. Table 1 shows that their total variance can reach 26.087%. It can also be learned from Table 3 that IP, a constituent of Factor 2, has the highest loading of 0.822, followed by AT and MH with loadings of 0.770 and 0.713, respectively. As mentioned earlier, Kaohsiung is a highly polluted industrial area. Besides being affected by pollution from mobile sources at the usual times, Kaohsiung is also affected by pollution from stationary sources, such as oil refineries, petrochemical plants, steel mills, and power plants. Higher concentrations of IP, AT, and MH among the VOCs emitted from stationary sources at specific times have been detected, in addition to those of the three VOCs that constitute Factor 1. Lower concentrations of IP, AT, and MH among the VOCs are detected at off-peak times since there are no emissions from most of the stationary sources. Currently, competent environmental authorities in the study area have formulated regulations governing the control of emissions from stationary sources [15,17]. However, stationary sources are distributed across the Pingtung Plain in the study area, whereby the dispersion of pollutants, such as malodorous substances, toxic chemical substances, or VOCs, from stationary sources is inevitable. When said pollutants react with sunlight, the photochemical reaction produces ozone [22]. The results of monitoring said reaction by Taiwan’s Environmental Protection Administration over the years show higher concentrations of IP, AT, and MH in the atmosphere. As the three VOCs that comprise of Factor 1 are emitted from mobile sources, they have a more significant impact on photochemical pollution. Thus, TU, TB, and IB are not parts of Factor 2. In other words, as Table 2 shows, TU, TB, and IB, among the constituents of Factor 1, have higher loadings of 0.879, 0.792, and 0.741, respectively. The results show that the three said VOCs can better represent and explain the pollution characteristics of Factor 1. Since TU, TB, and IB, among the constituents of Factor 2, generally have lower loadings of 0.211,  $-0.095$ , and 0.117, respectively, they cannot adequately represent and explain the pollution characteristics of Factor 2.

The above analysis results suggest that the main sources and emission behaviors of the IP, AT, and MH detected in the study area are related to the numerous stationary sources of air pollution. Hence, Factor 2 can be called “the factor of pollution from stationary sources.”

#### 3.4.3. Factor 3

As shown in Table 3, Factor 3 consists of four VOCs: EA, PP, NB, and NC. Table 1 shows that their total variance can reach 13.872%. It can also be learned from Table 3 that EA, a constituent of Factor 3, has the highest loading of 0.858, followed by PP, NB, and NC

with loadings of 0.780, 0.733, and 0.681, respectively. All four VOCs have straight-chain, single-bond alkane structures. Among them, EA, PP, and NB are the major components of natural gas, while the long-chain NC is generally found in liquefied petroleum gas. Currently, the oil refineries in Kaohsiung include Dalin Refinery in Xiaogang District, the liquefied natural gas terminal in Yongan District, and the No. 3 and No. 4 Naphtha Cracking Plants in the petrochemical industrial zone in Linyuan. These oil refineries and natural gas storage facilities separate the hydrocarbons of different boiling points via heating during the refining processes every day. Many VOCs are generated from said processes; the four VOCs that constitute Factor 3 are the first air pollutants generated from the processes. They are evidently the primary sources of photochemical pollution during the refining processes [23]. The liquefied natural gas terminal in Yongan District is currently the largest liquefied natural gas terminal. It can store up to 4.5 million tons of liquefied natural gas per year, providing power plants in Taiwan with natural gas all year round. The volume of natural gas to be used in Taiwan must reach 50% by 2025. As its storage volume increases, pollution caused by emissions of VOCs will inevitably increase. Moreover, there are several Naphtha Cracking Plants in Kaohsiung. Naphtha produced from crude oil during the refining process will decompose into VOCs with smaller molecules, such as ethylene, propylene, and butadiene [24]. Since crude oil contains a high concentration of long-chain NC, as well as that of EA, PP, and NB, its refining process can easily lead to the dispersion of VOCs and indirectly cause photochemical reactions.

The above analysis results suggest that the formation of EA, PP, NB, and NC in the atmosphere in the study area has to do with the liquefied natural gas terminal and Naphtha Cracking Plants in operation. The four VOCs are generated from the processes of refining the basic elements of non-renewable energy. Thus, Factor 3 can be called the “factor of pollution from energy sources.”

### 3.5. Results of Photochemical Pollution Characteristics Analyses—Cluster Analysis

A two-stage clustering process is used. First, hierarchical clustering is used for general clustering; then, the K-mean method is used to test the cluster numbers. Finally, five clusters are used to categorize the characteristics of the 10 VOC contaminants. Figure 2 shows the relationship between the clusters and the factors. Additionally, among the many VOC pollutants, some may share common characteristics. Understanding these characteristics will assist in analyzing the variations of relationship among the various photochemical pollution characteristics. The following characteristics can be derived from Table 4:

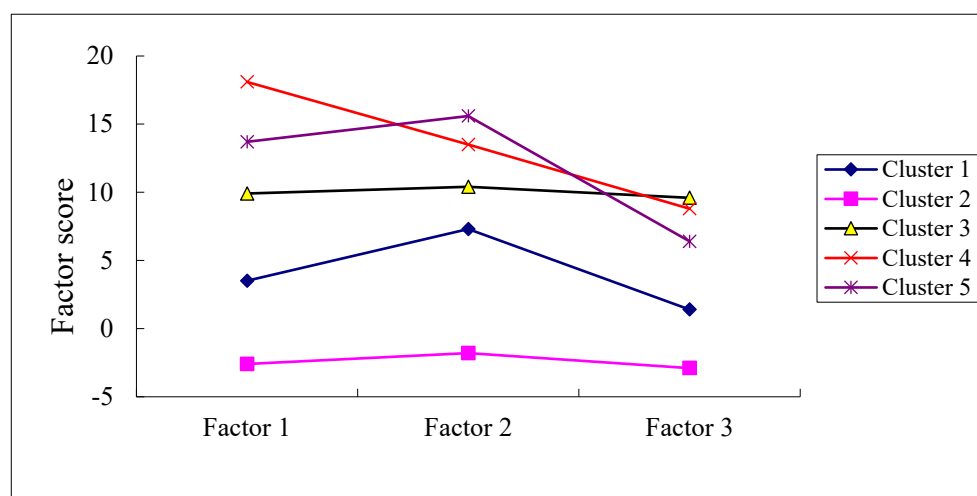


Figure 2. Factor scores of the five clusters.



**Table 4.** Average and maximum values for the VOCs in various clusters.

VOCs \ Clusters	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
TU (ppb)	4.89 0.24~50.4	1.39 0.12~36.7	5.97 0.71~70.9	9.78 1.12~112.4	6.55 0.56~82.8
TB (ppb)	5.24 0.45~46.1	2.59 0.08~41.7	3.49 0.43~53.5	8.76 2.08~79.3	4.77 0.72~65.9
IB (µg/m <sup>3</sup> )	4.02 0.44~25.8	1.71 0.09~20.3	3.46 0.71~63.6	6.74 0.96~80.5	3.92 0.53~66.6
IP (ppb)	6.02 0.23~21.8	3.08 0.11~13.7	7.03 0.45~28.6	7.49 1.26~36.2	8.58 1.78~42.1
AT (ppm)	4.08 0.09~22.0	2.97 0.06~16.3	6.85 0.31~35.9	7.64 0.96~23.8	9.23 2.56~50.5
MH (ppb)	3.57 0.11~6.8	2.10 0.05~3.8	4.38 0.24~15.1	3.58 0.32~26.1	6.10 1.37~43.0
EA (ppb)	4.54 0.36~29.9	2.87 0.11~39.8	6.43 0.39~47.8	4.48 1.24~45.6	7.32 2.44~60.8
PP (ppb)	2.25 0.17~14.2	1.73 0.12~30.3	5.06 0.69~99.1	4.87 0.90~97.4	4.06 0.88~89.6
NB (ppb)	2.32 0.31~21.4	1.04 0.13~30.1	2.58 0.47~89.2	3.47 0.40~102.3	3.16 0.53~71.3
NC (ppb)	2.07 0.19~7.5	1.09 0.09~6.8	3.76 0.28~33.6	4.03 0.32~18.1	3.58 0.47~29.1
Photochemical pollution conditions	General photochemical pollution	Mild photochemical pollution	Factor of moderate pollution from stationary sources and energy sources	Factor of moderate to severe pollution from mobile sources	Factor of moderate to severe pollution from stationary sources
The monitoring station where the VOCs are mostly detected	Chaozhou	Qiaotou and Chaozhou	Qiaotou and Linyuan	Xiaogang	Xiaogang and Linyuan
Number of samples (14,400 entries in total)	3186	7525	1457	1024	1208

(1) Cluster 1

As shown in Figure 2, Cluster 1 has the fourth highest scores on all the three factors: pollution from mobile sources, pollution from stationary sources, and pollution from energy sources. In other words, the concentrations of VOCs distributed across the study area are low in this cluster. Table 4 shows that the concentration values of the 10 VOCs in this cluster are only higher than those in Cluster 5, indicating an insignificant impact of photochemical reactions on this cluster. Moreover, VOCs in this cluster are generally detected in Chaozhou, Pingtung due to the fact that Chaozhou’s photochemical monitoring station is the only station among the four stations that is not located in Kaohsiung. Chaozhou is less affected by pollution from mobile sources since it has a fewer number of permanent residents; areas near it are nearly unaffected by pollution from stationary sources. However, it is worth noting that Chaozhou can be easily affected by a northeast monsoon in autumn and winter, and it is constrained by terrain. Air pollutants generated in Kaohsiung are transported to Pingtung by monsoons and settle in Chaozhou after being obstructed by the Central Mountain Range. Thus, there remain certain percentages of concentrations of VOCs in the atmosphere [25]. Overall, this cluster accounts for 22.1% of the total data entries (3186 out of 14,400 entries). Although the concentrations of VOCs in the atmosphere in this cluster are low, the impact of photochemical reactions on it is at a general level. This cluster is therefore classified as “the general photochemical pollution cluster.”

(2) Cluster 2

As shown in Figure 2, Cluster 2 has the lowest scores on all three factors. Moreover, its factor scores are much lower than those in the other four clusters. Table 4 shows that the concentration values of the 10 VOCs in this cluster are lower than those in other clusters,

and there is no sudden increase in the concentration values of VOCs, indicating a relatively insignificant impact of the atmospheric photochemical reactions on this cluster. In this cluster, the concentrations of VOCs reflected in the three factors are low; they are generally distributed across Qiaotou in Kaohsiung and Chaozhou in Pingtung. As mentioned earlier, areas near Chaozhou are less affected by pollution from stationary sources (i.e., there is no significant pollution from stationary sources) and that from mobile sources. Thus, the concentrations of VOCs detected in these areas are low. Although Qiaotou is close to the liquefied natural gas terminal in Yongan District, it is not significantly affected by pollution from energy sources. Moreover, significant stationary sources in Kaohsiung, such as Dalin Refinery, the No. 3 and No. 4 Naphtha Cracking Plants in the petrochemical industrial zone in Linyuan, and the Nanzih Export Processing Zone, are located south of Qiaotou. The air pollutants generated here are transported southward toward Pingtung by northeast monsoons in autumn and winter [3,24]. As Qiaotou is not a metropolitan area, pollution from mobile sources is not significant. Consequently, there is not a significant increase in the concentrations of VOCs detected in Qiaotou. This cluster accounts for 52.3% of the total data entries (7525 out of 14,400 entries). In other words, the study area is not affected by photochemical reactions for approximately more than half the year. This cluster can then be said to be the cluster least affected by photochemical reactions among the five clusters. It is therefore classified as “the mild photochemical pollution cluster.”

### (3) Cluster 3

As shown in Figure 2, Cluster 3 has the third highest scores on the factors of pollution from mobile sources and stationary sources, as well as the highest score on the factor of pollution from energy sources. Table 4 shows that the concentration values of all the VOCs in this cluster are higher than those in Clusters 1 and 2, especially that of PP, which constitutes part of the pollution from energy sources, and is the highest among all the clusters. This result indicates that areas where the VOCs in this cluster are detected can be more easily affected by the concentration of PP. As mentioned earlier, PP is a by-product generated from processing natural gas or refining crude oil. Moreover, when boilers are in operation, they often emit PP [24]. Table 4 shows that the atmospheric concentrations of PP can reach 99.1 ppb, while those of NB can reach 89.2 ppb, indicating a relatively significant impact of this factor of pollution from energy sources on this cluster. The concentrations of VOCs in this cluster are generally distributed across Qiaotou and Linyuan, which are located north and south of Kaohsiung, respectively. Both are on the outskirts of the Kaohsiung metropolitan area, and thus they are not as affected by the pollution from mobile sources as Xiaogang is. As such, the concentrations of the three VOCs that constitute Factor 1 are not abnormally high. To sum up, this cluster accounts for 10.1% of the total data entries (1457 out of 14,400 entries). Although pollution from energy sources is the main source of photochemical pollution for this cluster, its total variance is the lowest among the three factors. Its impact on photochemical pollution is not as significant as that of the other two factors, either. This cluster can thus be classified as “the cluster of moderate pollution from stationary sources and energy sources”.

### (4) Cluster 4

As shown in Figure 2, Cluster 4 has the highest score on the factor of pollution from mobile sources, as well as the second-highest scores on the factors of pollution from stationary sources and energy sources. Table 4 shows that the concentrations of TU, TB, and IB that constitute the factor of pollution from mobile sources can reach 112.4 ppb, 79.3 ppb, and 80.5 ppb, respectively, all of which are the highest among the five clusters. In addition to the factor of pollution from mobile sources, those of pollution from stationary sources and energy sources have an impact on this cluster, where the concentrations of VOCs are relatively high. The concentrations of VOCs in this cluster are distributed across Xiaogang due to the fact that it is within the Kaohsiung metropolitan area. There are fewer days that the concentration of ozone fails to comply with the air quality standards than those with an acceptable concentration of total suspended particulates (TSP) detected in the

study area [26,27]. In the Kaohsiung metropolitan area, there is a high ratio of using motor vehicles for commuting or travel purposes during holidays. The concentrations of TU, TB, and IB are higher among the pollutants emitted from mobile sources. Moreover, there is a high number of heavy industrial areas near Xiaogang, and Taiwan's largest international port, the Port of Kaohsiung, is in Kaohsiung. There are large vehicles traveling to and from said industrial areas and the Port of Kaohsiung at the usual times. The VOCs emitted from mobile sources and stationary sources can easily have photochemical reactions. They will form hydroxyl radicals after they are oxidized under ultraviolet irradiation in the atmosphere. Following that, they will react with nitric oxide (NO) to give rise to high concentrations of ozone and peroxyacetyl nitrates (PANs) [21]. In addition to pollution from mobile sources and stationary sources, pollution from energy sources, such as the PP, NB, and NC emitted from the steel mills, power plants, and oil refineries in Xiaogang, also has a significant impact on this cluster. This cluster accounts for 7.1% of the total data entries (1024 out of 14,400 entries). Most of the concentrations of VOCs in this cluster are detected in autumn and winter. It can thus be classified as "the cluster of moderate to severe pollution from mobile sources."

#### (5) Cluster 5

As shown in Figure 2, Cluster 5 has the highest score on the factor of pollution from stationary sources, as well as the second-highest score on the factor of pollution from mobile sources, among all the five clusters. Table 4 shows that the concentrations of IP, AT, MH, and EA that constitute the factor of pollution from stationary sources can reach 42.1 ppb, 50.5 ppb, 43.0 ppb, and 60.8 ppb, respectively, all of which are the highest among the five clusters. As mentioned in Cluster 4, Xiaogang in Kaohsiung has the highest number of heavy industrial areas and stationary sources in Taiwan. Coupled with the factor of pollution from mobile sources, which has the most significant impact on photochemical pollution, the concentrations of VOCs (especially that of aromatic hydrocarbons) from stationary sources detected in Xiaogang are generally higher than those detected in other areas. Moreover, there are two Naphtha Cracking Plants and a high number of stationary sources in Linyuan, which is located south of Xiaogang. In Cluster 3, we discussed the impact of the northeast monsoons on the VOCs detected in autumn and winter. Said VOCs are transported southward toward Pingtung by northeast monsoons [27]. However, the atmosphere in Linyuan is also affected by industrial production and pollution from mobile sources, and higher concentrations of VOCs are often detected in autumn and winter. This cluster accounts for 8.4% of the total data entries (1208 out of 14,400 entries), a proportion which is close to that of Cluster 4. To summarize, the primary source of photochemical reactions in this cluster is the factor of pollution from stationary sources. This cluster can thus be classified as "the cluster of moderate to severe pollution from stationary sources".

## 4. Conclusions

This study assessed the concentrations of 10 VOCs detected at four photochemical assessment monitoring stations that are part of the air quality monitoring network in the study area in southern Taiwan. The authors used factor and cluster analyses of the multivariate statistical analysis to explain all significant factors affecting ozone pollution levels and the mechanism for the formation of ozone in the study area. These influencing factors were properly classified into clusters that represent the photochemical reactions of different characteristics and photochemical pollution levels in communities within the study area. Ten VOCs of higher concentrations were selected from 54 ozone precursors as pollutants. The results of the factor analysis indicated three factors as the main factors affecting photochemical pollutants in the study area: pollution from mobile sources, pollution from stationary sources, and pollution from energy sources. K-means clustering was adopted for the cluster analysis to test a number of clusters. The conditions and levels of photochemical pollution in the study area can be classified into five clusters: "the general photochemical pollution cluster", "the mild photochemical pollution cluster", "the cluster of moderate pollution from stationary sources and energy sources", "the cluster of moderate to severe

pollution from mobile sources”, and “the cluster of moderate to severe pollution from stationary sources.” Among the five clusters, Cluster 4 is most severely affected by the three VOCs that constitute the factor of pollution from mobile sources, and thus it reflects the most significant level of photochemical reactions. Data on the 10 VOCs were collected over a period of one year and nine months because the analysis results are more stable when there are more types and samples of pollutants when performing a factor analysis, whereby the validity and reliability of the factor analysis results can be effectively improved. The findings of this study can effectively create a classification of the characteristics of photochemical pollution and a mechanism for the formation of photochemical pollutants, which can serve as references for environmental agencies when implementing air quality management strategies.

**Author Contributions:** The first author E.M.-Y.W. formulates the research of direction of this study clearly, and is responsible for defining the topic and organizing the research. Besides, he collects and compiles data of four photochemical assessment monitoring stations in the Kaohsiung-Pingtung area. S.-L.K. mainly carries out the execution of SPSS Statistics 26 and explains all significant factors affecting ozone pollution levels and the mechanism for the formation of ozone. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Data are available upon request.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Karle, M.M.; Mahmud, S.; Sakai, R.; Fitzgerald, R.M. Investigation of the successive ozone episodes in the El Paso–Juarez region in the summer of 2017. *Atmosphere* **2010**, *11*, 532. [\[CrossRef\]](#)
2. Fujita, E.M.; Campbell, D.E.; Stockwell, W.R.; Lawson, D.R. Past and future ozone trends in California’s south coast air basin: Reconciliation of ambient measurements with past and projected emission inventories. *J. Air Waste Manag. Assoc.* **2013**, *63*, 54–69. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Kuo, H.W.; Wei, H.C.; Liu, C.S. Exposure to 16 volatile organic compounds while commuting in Taichung Taiwan. *Atmos. Environ.* **2000**, *34*, 3331–3336. [\[CrossRef\]](#)
4. Kaipara, R.; Rajakumar, B. Oxidative degradation kinetics and mechanism of two biodiesel constituents (ethyl-2-methyl propionate (E2MP) and ethyl-2,2-dimethyl propionate (E22DMP)) initiated by Cl atoms. *ACS Earth Space Chem.* **2020**, *4*, 142–156. [\[CrossRef\]](#)
5. Balan, R.C.; Rajakumar, B. Photo oxidation reaction kinetics of ethyl propionate with Cl atom and formation of propionic acid. *J. Phys. Chem. A* **2018**, *122*, 8274–8285.
6. Chou, T.G. *The Research of Air Quality Management in Fine Particulate Matters*; Central Office of Administration, Academia Sinica Research weekly 1276; Central Office of Administration, Academia Sinica: Taipei, Taiwan, 2010. (In Chinese)
7. Chatzidiakou, L.; Mumovic, D.; Summerfield, A.J.; Hong, S.M.; Altamirano-Medina, H. A victorian school and a low carbon designed school: Comparison of indoor air quality, energy performance, and student health. *Indoor Built Environ.* **2014**, *23*, 417–432. [\[CrossRef\]](#)
8. Pegas, P.N.; Alves, C.A.; Evtyugina, M.G.; Nunes, T.; Cerqueira, M.; Franchi, M.; Pio, C.A.; Almeida, S.M.; Verde, S.C.; Freitas, M.C. Seasonal evaluation of outdoor/indoor air quality in primary schools in Lisbon. *J. Environ. Monit.* **2011**, *13*, 657–667. [\[CrossRef\]](#)
9. Gupta, P.; Rajakumar, B. Reaction kinetics of a series of alkanes with ClO and BrO radicals: A theoretical study. *Int. J. Chem. Kinet.* **2020**, *53*, 189–206. [\[CrossRef\]](#)
10. Hsu, C.H.; Cheng, F.Y. Synoptic weather patterns and associated air pollution in Taiwan. *Aerosol Air Qual. Res.* **2019**, *19*, 1139–1151. [\[CrossRef\]](#)
11. Gouriou, F.; Morin, J.P.; Weill, M.E. On-road measurements of particle number concentrations and size distributions in urban and tunnel environments. *Atmos. Environ.* **2004**, *38*, 2831–2840. [\[CrossRef\]](#)
12. Baidya, B.; Chandra, A.K. Theoretical investigation on atmospheric chemistry of (CF<sub>3</sub>)<sub>2</sub>C=CH<sub>2</sub>: Reaction with OH radical. *Chem. Phys. Lett.* **2020**, *749*, 137409. [\[CrossRef\]](#)
13. Rivela, C.B.; Tovar, C.M.; Teruel, M.A.; Barnes, I.; Wiesen, P.; Blanco, M.B. CFCs replacements: Reactivity and atmospheric lifetimes of a series of Hydrofluoroolefins towards OH radicals and Cl atoms. *Chem. Phys. Lett.* **2019**, *714*, 190–196. [\[CrossRef\]](#)

14. Bengraïne, K.; Marhaba, T.F. Using principal component analysis to monitor spatial and temporal changes in water quality. *J. Hazard. Mater. B* **2003**, *100*, 179–195. [[CrossRef](#)]
15. Arslan, O.A. GIS-based spatial-multivariate statistical analysis of water quality data in the Porsuk river, Turkey. *Water Qual. Res. J. Can.* **2009**, *44*, 279–293. [[CrossRef](#)]
16. Mendiguchia, C.; Moreno, C.; Galindo-Riano, D.M.; Garcia-Vargas, M. Using chemometric tools to assess anthropogenic effects in river water: A case study Guadalquivir River (Spain). *Anal. Chim. Acta* **2004**, *515*, 143–149. [[CrossRef](#)]
17. Singh, K.P.; 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*, 3980–3992. [[CrossRef](#)]
18. Cheng, W.; Chen, Y.; Zhang, J.; Lyons, T.J.; Pai, J.; Chang, S. Comparison of revised air quality index with the PSI and AQI indices. *Sci. Total Environ.* **2007**, *382*, 191–198. [[CrossRef](#)]
19. Kuo, S.L.; Wu, E.M.Y. Highway tunnel air quality assessment using multivariate statistical classification on factor, cluster, and discriminant analysis. *Int. J. Eng. Technol.* **2018**, *7*, 287–291.
20. Tobiszewski, M.; Tsakovski, S.; Simeonov, V.; Namieśnik, J. Surface water quality assessment by the use of combination of multivariate statistical classification and expert information. *Chemosphere* **2010**, *80*, 740–746. [[CrossRef](#)]
21. Cheng, J.H.; Hsieh, M.J.; Chen, K.S. Characteristics and source apportionment of ambient volatile organic compounds in a science park in central Taiwan. *Aerosol Air Qual. Resour.* **2016**, *16*, 221–229. [[CrossRef](#)]
22. Baidya, B.; Lily, M.; Chandra, A.K. Theoretical insight into the kinetics of H-abstraction reaction of CHF<sub>2</sub>CH<sub>2</sub>OH with OH radical atmospheric lifetime and global warming potential. *ChemistrySelect* **2018**, *3*, 6136–6144. [[CrossRef](#)]
23. Tovar, C.M.; Blanco, M.B.; Barnes, I.; Wiesen, P.; Teruel, M.A. Gas-phase reactivity study of a series of hydrofluoroolefins (HFOs) toward OH radicals and Cl atoms at atmospheric pressure and 298K. *Atmos. Environ.* **2014**, *88*, 107–114. [[CrossRef](#)]
24. Chen, W.K.; Li, T.D. *The Establishment of Aerosol Control Mode Based on the Photochemical Reaction Mechanism*; NSC-1999-EPA-Z-231-001; National Science Council, Executive Yuan: Taipei, Taiwan, 1999. (In Chinese)
25. Shi, X.; Qiu, X.; Cheng, Z.; Chen, Q.; Rudich, Y.; Zhu, T. Isomeric identification of particle-phase organic nitrates through gas chromatography and time-of-flight mass spectrometry coupled with an electron capture negative ionization source. *Environ. Sci. Technol.* **2020**, *54*, 707–713. [[CrossRef](#)] [[PubMed](#)]
26. Kumar, U.; Prakash, A.; Jain, V.K. A multivariate time series approach to study the interdependence among O<sub>3</sub>, NO<sub>x</sub> and VOC<sub>5</sub> in ambient urban atmosphere. *Environ. Model. Assess.* **2009**, *14*, 631–643. [[CrossRef](#)]
27. Maurer, M.; Klemm, O.; Lokys, H.L.; Lin, N.H. Trends of fog and visibility in Taiwan: Climate change or air quality improvement? *Aerosol Air Qual. Res.* **2019**, *19*, 896–910. [[CrossRef](#)]