applied sciences

MDPI

Article

Investigation of VOC Series Collected in a Refinery and Their Classification Based on Statistical Features

Alina Bărbulescu [1], Sebastian-Barbu Barbeş [2], Lucica Barbeş [3,4,*] and Cristian Ștefan Dumitriu [5,*]

[1] Department of Civil Engineering, Transilvania University of Brașov, 5 Turnului Str., 500152 Braşov, Romania; alina.barbulescu@unitbv.ro
[2] Doctoral School of Civil Engineering, Technical University of Civil Engineering of Bucharest, 122-124 Lacul Tei Bvd., 020396 Bucharest, Romania; sebastian-barbu.barbes@phd.utcb.ro
[3] Department of Chemistry and Chemical Engineering, Ovidius University of Constanța, 124 Mamaia Bvd., 900527 Constanta, Romania
[4] Doctoral School of Biotechnical Systems Engineering, Politehnica University of Bucharest, 313 Splaiul Independenţei, 060042 Bucharest, Romania
[5] Faculty of Mechanical Engineering and Robotics in Constructions, Technical University of Civil Engineering Bucharest, 59 Calea Plevnei, 010223 Bucharest, Romania
* Correspondence: lucille.barbes2020@gmail.com (L.B.); cristian.dumitriu@utcb.ro (C.Ș.D.)

Featured Application: The results of the article can be applied to determine the zones were the maintenance works should be applied on the tank farm.

Abstract: In the context of the increased pollution from different sources and its significant negative effect on the population's health and environment**,** the article presents a comprehensive analysis of the data series formed by the concentrations of the volatile organic compounds (VOCs) collected in three zones—storage areas in the reservoir park—of a refinery complex in Romania during the maintenance period. Statistical analyses, including parametric and nonparametric tests, were performed to assess the correlation between the studied series and to group them based on some common features. The series were clustered using the raw data, and the series features were extracted after the statistical analysis. The results indicate that the series are not correlated and do not follow the same distribution even though the study zone is not large. The sites' classification based on statistical features is shown to be more relevant from the viewpoint of the emissions level than that provided using the raw series. The Principal Component Analysis (PCA) indicates that the features with the highest contribution on the first two components are maximum, standard deviation, autocorrelation, and partial autocorrelation for Zone 1; average, maximum, minimum, and partial autocorrelation for Zone 2; and skewness, average, maximum, and standard deviation for Zone 3. The study's novelty is two-fold. First, it provides the results of the study performed during the maintenance period of the storage tanks, which was insufficiently investigated in the literature. Secondly, since complete data series are not generally available to the large public, clustering them based on their features provides a clear image of pollution levels and the sites where actions should be taken to reduce it. This investigation offers essential insights that can serve as a background for developing effective air pollutant monitoring strategies and mitigation measures by understanding the emission patterns and identifying the factors that influence VOC levels during the maintenance of storage tanks for highly volatile petroleum products.

Academic Editor: Athanasia Koliadima

Received: 18 November 2024
Revised: 13 December 2024
Accepted: 17 December 2024
Published: 19 December 2024

Citation: Bărbulescu, A.; Barbeş, S.B.; Barbeş, L.; Dumitriu, C.Ş. Investigation of VOC Series Collected in a Refinery and Their Classification Based on Statistical Features. *Appl. Sci.* **2024**, *14*, 11921. https://doi.org/10.3390/app142411921

Keywords: VOCs; statistical analysis; features; classification

1. Introduction

The study of air quality on industrial platforms has become increasingly important due to phenomena that directly manifest through the zonal concentration of pollutants in some regions or indirectly through phenomena such as acid rain, ozone layer depletion, and photochemical smog. Various anthropogenic and natural sources emit Volatile Organic Components (VOCs). Given their photochemical property, VOC may react with nitrogen oxides from the atmosphere, resulting in secondary products that may injure the population's health (Evans et al., 2002) [1]. Benzene and 1–3-butadiene are highly toxic [2–4]. They frequently affect the health of workers from the refineries and the population in the settlements nearby [5–7]. Therefore, the VOC series study is essential as an early warning instrument for assessing air quality and monitoring other pollutants in the same area [8], maintaining the air quality and avoiding negative impacts on people's health, especially when the refineries are situated in urban zones.

Different software has been developed to facilitate modeling various categories of pollutants. For example, the US-EPA has developed comprehensive atmospheric dispersion modeling packages incorporating applications like AERMOD (AMS/EPA Regulatory Model) or ISCST (Industrial Source Complex—Short Term) into a single interface. These models are based on the Gaussian normal distribution, which the pollutants' series does not always follow. Therefore, the results show [9] the overestimation of the pollutants' concentrations. Still, AERMOD proved to be a valuable tool for modeling various pollutant series. For example, Saikomol et al. [10] explored the spatial dispersion of VOCs in a petroleum products tank farm of a crude oil refining complex in eastern Thailand in various scenarios, including the presence of a vapor recovery unit (VRU), double sealing of the storage tank, and modifications to the external floating roof of the tank. In a study by Karbasi et al. [11], the dispersion of benzene was evaluated across four seasons at a petroleum waste storage site on Kharg Island (Iran). The authors employed TANKS 4.0.9 (US EPA) benzene's atmospheric emission rates and simulated benzene series dispersion over an area of 10 × 10 km^2 with AERMOD.

Digital GIS representations can be a powerful tool in air quality management. For instance, Milazzo [9] used them to show the distribution of VOCs during loading ships with petroleum products. This visualization provided a clear understanding of the problem and facilitated the proposal of various measures for reducing benzene emissions.

Shahbazi et al. [12] undertook a comprehensive study of VOC emissions in Tehran in 2013, utilizing the IVE (International Vehicle Emissions) model. This model considers various factors influencing emissions. Similarly, Shie and Chan [13] performed a spatio-temporal analysis of the atmospheric pollutant immissions produced during and after a fire at a petrochemical complex in Taiwan, employing the Back Trajectory (BT) model and Surfer software 9.0.

Multiple linear and nonlinear regressions and ANNs proved useful for Hong et al.'s [14] research on the accumulation of BTEX on road surfaces encompassing typical commercial, industrial, and residential lands. Liu and Li [15] applied the COMIS model [16] to the benzene emission rates series from an apartment in China.

Zapata-Marin [17] used a combination of geostatistical methods and land-use regression to forecast the VOC pollution in Montreal, demonstrating the practicality and versatility of these techniques. Similarly, Jung et al. [18] employed spatial interpolation methods to emphasize the population exposure to different atmospheric pollutants, including VOCs, showcasing the broad applicability of these methods.

During the last period, many articles discussed the pollution level in Romania [19–26]. However, only a few addressed the pollution in urban or industrial zones [5,27,28]. The present study aims to evaluate the degree of pollution with VOCs in the area of an oil refinery in the northern part of the Romanian Littoral. As explained in the following section, the study area's strategic importance, added to by the presence of hazardous substances regulated by Romanian Law No. 59/2016, emphasizes the relevance of understanding emissions in this context. The research offers important insights into VOC emissions in a refinery during maintenance in the context of the environmental implications and impact on the workers and people living in the Năvodari and Mamaia cities and neighboring villages.

An extended statistical analysis of 41 series of pollutants recorded in three zones of the storage tanks during a maintenance period (one month) is performed. The series' correlation and similarities are assessed to understand the series characteristics, which are the basis of grouping the observation sites into clusters. Considering ten statistical futures, the series from each zone are clustered using the k-means algorithm. We show that this procedure allows a more effective assignation of each series in the groups than when using the raw data series for the same purpose. Its advantage lies in its adaptability to various case studies. It emphasizes that knowledge of some basic statistics allows a fast analysis of the pollution level during a specific period, providing valuable insights and information. Our search of the statistical literature did not return results on similar attempts. Moreover, our study aligns with or advances current trends in air quality management in industrial settings by:

- *Involvement of VOCs in Global Environmental Challenges.* VOC emissions contribute significantly to the formation of tropospheric ozone and other particles present in the atmosphere, affecting human health and ecosystems. Our study makes an important contribution by highlighting emissions from a critical operational phase—the verification period for high-volatility petroleum storage tanks—which is often overlooked in environmental reports.
- *Alignment with the General International/National Regulatory Framework.* The analysis of atmospheric pollutant concentrations with respect to the EU and Romania's standards is based on measurements taken at fixed sampling points. Our research provides an analytical framework adapted to meet the requirements of European, national, and local regulations. Since emission/immision monitoring differs by parameters of interest, the limit concentrations, and observation frequencies, the regulations in the field are different. The reference documents taken into account, which present the codes and regulations required in the design of both petroleum storage tanks under atmospheric pressure and those under pressure are the Decision (EU) 2017/1757 of the Council of 17 July 2017, regarding the acceptance, on behalf of the European Union, of an amendment to the 1999 Protocol to the 1979 Convention on Long-Range Transboundary Air Pollution to reduce the degree of acidification, eutrophication, and the level of tropospheric ozone [29] and Law no. 264 of 20 December 2017, regarding the establishment of technical requirements to limit emissions of volatile organic compounds (VOCs) resulting from gasoline storage and distribution from terminals to gas stations, as well as during vehicle refueling at gas stations [30]. Identifying critical areas and classifying data series based on their statistical characteristics facilitates decision-making for compliance and pollution reduction in industrial areas.
- *Operational Impact*. Our study's conclusions will provide refineries with a practical tool, allowing them to prioritize interventions in areas with the highest emissions and optimize the planning of checks to reduce their environmental impact.

2. Study Area

The study site (Figure 1) is situated north of Constanța (Romania), at the boundary of Năvodari town, approximately 100 m from the shore of the Black Sea. The total area of the industrial platform is 2 km², and the experimental zone (the tank park for storing petroleum products) is approximately 0.5 km². The sampling area is located at latitude 44°20′12″ N–44°20V30″ N and longitude 28°38′56″ E–28°39′6″ E and was divided into three distinct zones (1, 2, and 3, in Figure 1). The characteristics of the experimental zones are presented in Table 1.

A moderate thermal regime characterizes the climate of the experimental area. During summer days, the vertical thermal stratification of the air is more stable, resulting in descending vertical currents that prevent cloud formation.

Figure 1. Sampling area. The dots represents the sampling points in Zones 1, 2, and 3.

Table 1. Characteristics of the experimental area.

Zone	Access Road	No. of Sampling Points	No. of Tanks Under Repair	Tank Capacity (m³)	Transported Product
1	R20	15	1	5000 (×4)	gasoline/rafined product
2	R4	14	1	10,000 (×4) and 3500 (×2)	gasoline
3	R9	12	2	2000 (×2) and 1000 (×2)	toluene and gasoline

Thus, cloudiness is reduced, and the duration of exposure to high temperatures is longer than in the rest of Romania. The recorded air humidity annual average is approximately 80%; the relative minimum frequently recorded in July is 70%, and the relative maximum recorded in December is 90%. The annual number of days with 80% humidity is 129 days. Precipitation has a torrential character in the warm half of the year and alternates with long periods of drought. Barometric pressure values fall within the range: a minimum of 978.1 mbar, a multiannual average of 1013.3 mbar, and a maximum of 1056.4 mbar. The absolute monthly minimum and maximum temperature values recorded in January are −24.7 °C and +18 °C, respectively, and in February are −25.0 °C and +24.5 °C, respectively. Because clear weather predominates in the experimental location area and there is a relatively high frequency of tropical and continental air invasions, the daily maximum air temperature exceeds 25 °C more than 60 days per year. Days with a maximum temperature higher than 20 °C are frequent during summer, especially in July and August

when their monthly average number exceeds 20 days. In spring and autumn, the number of days with a maximum temperature higher than 20 °C is lower. During winter, temperatures lower than −25 °C have not been observed. Solar radiation depends on cloudiness, with heliographic data establishing that the annual duration of sunshine is 2330 h. The yearly average wind speed is approximately 2–5 m/s with a frequency of 40% and 6–10 m/s with a frequency of 30%.

3. Materials and Methods

3.1. Data Collection

The VOC concentrations data series were collected during three experimental campaigns from 25 September to 24 October 2023, 20 November–19 December 2023, and 18 January–17 February 2024, corresponding to the revision periods of some tanks in zones 1–3 from Figure 1. The measurements were done at a high of 1.4 m from the ground every 36 h. The first measurement was done on the first day at 7 a.m., the second one on the second day at 7 p.m., and so on, due to the restrictions imposed by the maintenance schedule. We worked with the average series from the three campaigns to ensure the data series' reliability. Therefore, there are 20 values for each measurement point in each zone. Thus, the datasets comprise 15 series $\times$ 20 = 300 values in Zone 1, 14 series $\times$ 20 = 280 values in Zone 2, and 3 by 13 series $\times$ 20 = 260 values in Zone.

In monitoring VOC emissions from the tank park designated for storing highly volatile petroleum products, both FID (Flame Ionization Detectors) and PID (Photoionization Detectors) analyzers were used as follows:

Portable Device TVA 1000B, manufactured by Thermo Scientific (Waltham, MA, USA), contains a dual detection system (FID and PID). The technical specifications of the measurement equipment are measurement range (0.5–2000 ppm isobutylene (PID)/0.5–5000 ppm methane (FID)); response time (3.5 s); accuracy 2.5 ppm (FID and PID); linearity 0.5–500 ppm isobutylene (PID)/0.5–10,000 ppm methane (FID); repeatability (±1% (PID); ±2% (FID)); flow rate 1000 mL/min under environmental conditions; operating time—8 h; standard calibration—FID (methane)/PID (isobutylene).

Portable Device Dräger Multi-PID 2, manufacturer Ion-Science (Fowlmere, UK), with the following technical specifications: minimum resolution (1 ppb); range of 0–20,000 ppm; response time <2 s; standard calibration 100 ppm with customizable calibration capability using isobutylene; flow rate—220 mL/min; PID lamp of 10.6 eV (standard) and two other lamps of 11.7 eV available.

At the beginning of each working day, the FID is calibrated with methane, and the PID is calibrated with isobutylene to ensure the manufacturer's specified accuracy for the TVA 1000B portable device.

3.2. Methodology

The study's stages are presented below.

I. Compute the data series' basic statistics to assess the series' characteristics.

II. Perform the following statistical tests to determine the series properties and assess their similarity/dissimilarity, as presented below.

- Test the series normality using the Shapiro–Wilk [31] and the Q–Q plot to cross-validate the results [32]. The Q–Q plot is obtained by plotting the estimated quantiles of a distribution against those of another distribution (generally considered as a reference, in our case, Gaussian).
- Test the series homoskedasticity by the Levene test [33].

- Drawing the boxplots to assess the outliers' existence. An outlier is a value of the data series outside the interval [Q_1 −1.5 × *IQR*, Q_3 +1.5 × IQR], where Q_1 (Q_3) is the lower (upper) quartile is, and IQR is the interquartile range = $Q_3 - Q_1$.
- In this study, we did not eliminate the outliers. Excluding outliers would not be beneficial for detecting the areas with more accentuated pollution. As presented in the analysis, the outliers' existence is one of the features used in the second phase of clustering.
- Compute the autocorrelation function (ACF) and draw the correlogram to see the variation in the correlations between couples of the series values at different lags.
- For a time series (X_t), ACF is defined by:

$$\boldsymbol{\rho_k} = \frac{\boldsymbol{Cov(X_t, X_{t+k})}}{\sqrt{\boldsymbol{Var(X_t)Var(X_{t+k})}}} \quad (1)$$

where $Cov(X_t, X_{t+k})$ is the covariance of X_t and X_{t+k}, and $Var(X_t)$ ($Var(X_{t+k})$) is the variance of X_t (X_{t+k}).

- Compute the partial autocorrelation function (PACF), which indicates the correlation between X_t and X_{t+k} considers intermediary lagged values, in contrast to the ACF, which takes into account all the intermediary values.

III. Study the relationships between series (and their characteristics) as a group.

1. Evaluate the correlations between the series in each zone. The first approach involved building the correlation matrices, whose elements are the Pearson correlation coefficients of couples of series. The second one was checking the multivariate independence of the series in each group (not pairwise, as in the correlation matrix), using the distance correlation t-test of independence [34]. The advantage of the last procedure is that it uses a modified distance that makes it robust compared to other tests based on second-order moments.
2. Test the hypothesis that all series in a zone come from the same continuous distribution against the hypothesis that at least one is issued from another distribution using the Anderson–Darling test implemented in the 'kSamples' package in R [35].
3. Use Mood's median test [36] to check the hypothesis that the medians of the series in each group are equal against the alternative that at least two differ.
4. Compare the variances of the collected samples utilizing the Levene test (with medians instead of means) and the Fligner–Killeen [37] nonparametric test (based on ranks). The null hypothesis is that the variances of all samples are equal, and the alternative is that at least two variances differ. The above tests were utilized due to their robustness to the normality hypothesis violation.

IV. Group the data series in clusters using the k-means algorithm [38,39]. The algorithm was employed to cluster the raw data series, on the one hand, and the series of features determined from statistical analysis, on the other hand, after normalizing the datasets.

Finding the optimal number of clusters (*k*) is essential for robust clustering. When selecting *k*, different methods can be used, like those implemented in the NBClust package in R 4.4.2 [40]. They are based on various mathematical approaches, which are not equivalent to each other, so the provided optimal number of clusters differs. This is why the maximum criterion is utilized to select the best *k*. Many authors use the most known method for selecting k—the silhouette method. Once *k* is selected, the discussion of cluster stability strengthens the choice, with the connectivity criteria being put in second place or

neglected, given that it implicitly intervenes, to a certain extent, in grouping the series into clusters.

A criterion for selecting between two clusterings is the maximum separability between clusters, represented as the ratio BSS/TSS (the sum of squared distances inside the clusters over the sum of squared distances between the clusters). The higher the ratio, the better the clustering is.

Connectivity is a local measure of how well the points within a cluster are related to each other, but it does not capture stability across multiple clustering runs. It helps evaluate intra-cluster compactness rather than the overall consistency of the clustering process. Stability indicators are broader measures designed to quantify how stable the clustering results are when the algorithm is applied multiple times. For example, using bootstrapping, stability indicators assess the variation in cluster assignments when the data are perturbed (through resampling, for instance). In the case of k-means clustering, stability indicators are particularly helpful when the data are noisy or when the choice of the number of clusters is uncertain. Therefore, stability is the criterion that should be given priority after detecting the optimal number of clusters when various methods provide different values of *k*.

This study assessed the internal clustering connectivity using the Silhouette width, Dunn index, and connectivity. The first coefficient takes values between −1 and 1 and should be maximized. The other two are positive and should be minimized—Dunn or maximized—connectivity. The clusters' stability was evaluated by the mean distance between means (ADM), average distance (AD), average proportion of non-overlap (APN), the figure of merit (FOM) [41,42], and average Jaggard measure (AJ). The lower the ADM, APM, and FOM, the better the stability. AJ values above 0.85 indicate high stability of a particular cluster, between 0.6 and 0.85 its stability, and less than 0.6 its instability [43].

Given the various computational methodologies, discordances in the clustering validations—stability versus connectivity—are likely to appear. No selection criterion is perfect. Thus, the scientist must balance the criterion for choosing k with the cluster's stability and connectivity. Therefore, in this study, we tried to strike a balance between all the criteria when choosing clustering.

The k-means was chosen because it has the following advantages against the hierarchical clustering [44]:

- It is relatively simple and easy to implement;
- Has a lower computational complexity, which allows it to handle datasets more efficiently;
- Has a higher computational efficiency and scalability accuracy;
- Mean or median can be used as a cluster center, offering flexibility;
- It is less sensitive to outliers.

Compared to Gaussian mixture models [45], the k-means algorithm is less computationally intensive [46]. Moreover, GMMs can overfit the data, capturing noise rather than the underlying distribution.

Since the Anderson–Darling test rejected the series normality in most cases and we found that various data series come from some kinds of simple distributions (as shown in Results), the Gaussian mixture was not preferred for clustering.

V. Determine the principal components (PCs) that significantly influence clustering based on the statistical features. The Principal Component Analysis (PCA) is used to reduce a dataset's dimensionality "while retaining as much as possible of the variation present in the dataset. This is achieved by transforming to a new set of variables, the principal components (PCs), which are uncorrelated and ordered so that the first few retain most of the variation present in all of the original variables" [47].

PCA was performed to determine the features with the most significant contributions to the new axes (dimensions, PCs). The features with the highest contributions on the PCs are the most important in characterizing the series variability. When data series are unavailable, and only the 'basic statistics' are available, knowing which information should be prioritized for fast analysis is essential. In our study, the biplot provided by PCA can emphasize the essential features that help characterize the pollution level and the points with the highest contribution to pollution. The results will show where it is necessary to take action to eliminate the contamination.

To avoid repetitions in the text, we shall refer to a point from a zone by Z*i*M*j*, where *i* is the zone number (*i* = 1, 2, 3) and *j* takes values from 1 to 15 (for Zone 1), 1 to 14 (for Zone 2), and 1 to 12 (for Zone 3). For example, Z1M7 is point 7 in Zone 1.

4. Results

I. The charts of the series collected in Zone 1 (2 and 3) are contained in Figure 2. In Zone 1, 13 points were located around four reservoirs. These facilities were selected for maintenance work to check for seal integrity. The highest concentrations of pollutants were recorded at the extreme edges of the monitored area, predominantly at two points located along a main access road (R20) leading to the reservoirs. The distribution suggests that the pollutants were predominantly dispersed in the direction of this access road, possibly due to existing road traffic, as well as direct exposure from emission sources to the two monitored points, which had a wider opening, facilitating the spread of VOCs without natural or artificial barriers. The potential sources of these pollutants could be direct exposure to emission sources and road traffic.

Data analysis indicates that there were no significant accumulations in the points of Zone 1, which are located along the sides or at the intersections of the diagonals of the rectangle circumscribed around the circle formed by the centers of the four monitored reservoirs.

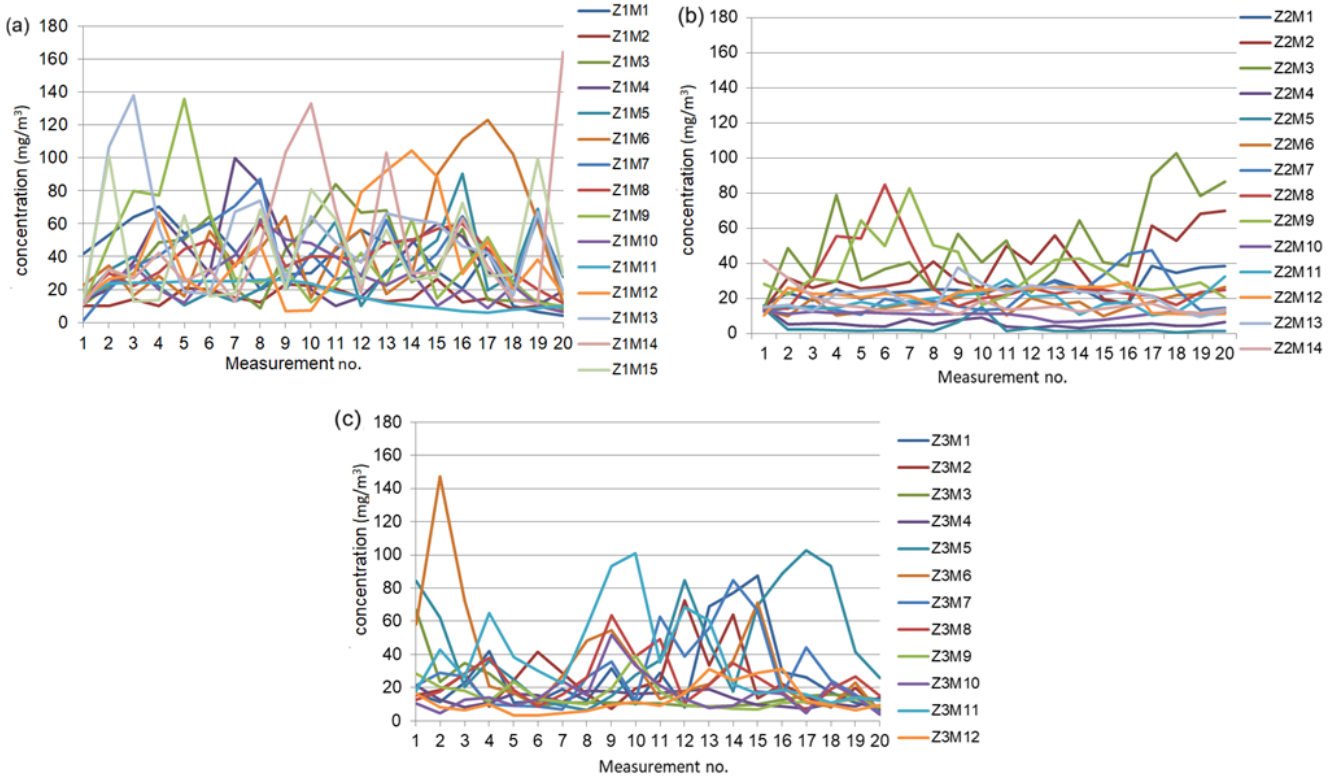

Figure 2. Series recorded at the locations from (**a**) Zone 1, (**b**) Zone 2, and (**c**) Zone 3.

In Zone 2, among the 14 measurement points, nine were selected near six reservoirs under maintenance, and 5 points were positioned along the main access road (R4) to the targeted reservoirs. The VOC concentration values exhibited distinct characteristics depending on the location. At the measurement points within the monitored area adjacent to the access road, pollutant concentrations (minimum, maximum, and average) were comparable, indicating a relatively uniform dispersion of VOCs. The result suggests that emissions are evenly distributed near the reservoirs, without significant variations between points on either side of the R4 access road. However, the monitoring points on the access road recorded slightly higher pollutant concentrations. These results can be explained by the accumulation of emissions along the road, which acts as a preferential dispersion pathway, amplifying local VOC concentrations. This observation is supported by the fact that the highest pollutant concentration values were detected at the measurement points located at the ends of the access road. This observation indicates an accumulation of VOC emissions in these areas, where dispersion is less effective, possibly due to physical barriers created by industrial structures (e.g., dikes), various operating systems used, such as petroleum product pumping stations, or determined by the prevailing wind direction. The statistical analysis results may reflect the dynamics of pollutant dispersion, indicating that peripheral points are more exposed to high concentrations due to the accumulation mechanism, local conditions created by air currents, and the site's topography.

In Zone 3, the highest pollutant concentrations were found on the access roads (R4 and R9), leading to the monitoring of four other reservoirs undergoing maintenance. These reservoirs are located near the R4 access road in Zone 2 and a second access road (R9) in Zone 3, which are perpendicular to each other. In zone 3, it was observed that the highest pollutant concentrations were recorded at a point near a reservoir that did not have a secondary seal on its cap. The lack of sealing on the reservoir led to uncontrolled VOC emissions, significantly increasing local pollutant concentrations. The seal lack finding underscores the urgent need for effective sealing systems to prevent unchecked pollutant emissions, especially during maintenance periods when the risk of emission leaks is higher.

Tables S1–S3 from the Supplementary Material contain the latitudes and longitudes of the observation points and the values of the basic statistics. The series of basic statistics are depicted in Figure S1.

The following variation intervals of the values are determined in the study zones (1, 2, and 3, respectively):

- For the minimum (mg/m^3): [1.2, 12.69], [0.62, 15.75], and [3.96, 11.15];
- For maximum (mg/m^3): [25.98, 164.53], [12.98, 102.47], and [21.24, 147.42];
- For the average (mg/m^3): [14.95, 52.29], [3.21, 50.56], and [12.93, 45.33];
- For the standard deviation (mg/m^3): [5.48, 43.39], [1.85, 24.82], and [4.16, 33.37];
- For the skewness coefficient: [−0.37, 1.48], [−1.33, 2.53], and [0.03, 2.92];
- For the kurtosis: [−1.86, 1.51], [−1.07, 7.41], and [−1.25, 5.85].

The limits of the intervals to which the standard deviations belong vary significantly. The same is true for the skewness and kurtosis of the series recorded in different zones. We notice right- and left-skewed series in Zone 1 and mostly (almost) symmetrical and right-skewed series in the other zones. Maxima varied in large limits, with the highest amplitude in Zone 3 and the lowest in Zone 2. The maps of the spatial distributions of minimum, maximum, and average concentrations in Zone 1 (2 and 3) are presented in Figure 3 (Figures S2 and S3). The maps were drawn considering the results from all measurement campaigns, averaging them, as explained in Methodology. Given the relatively low height at which the measurement was done and the walls around the tank area, the influence of wind is not significant.

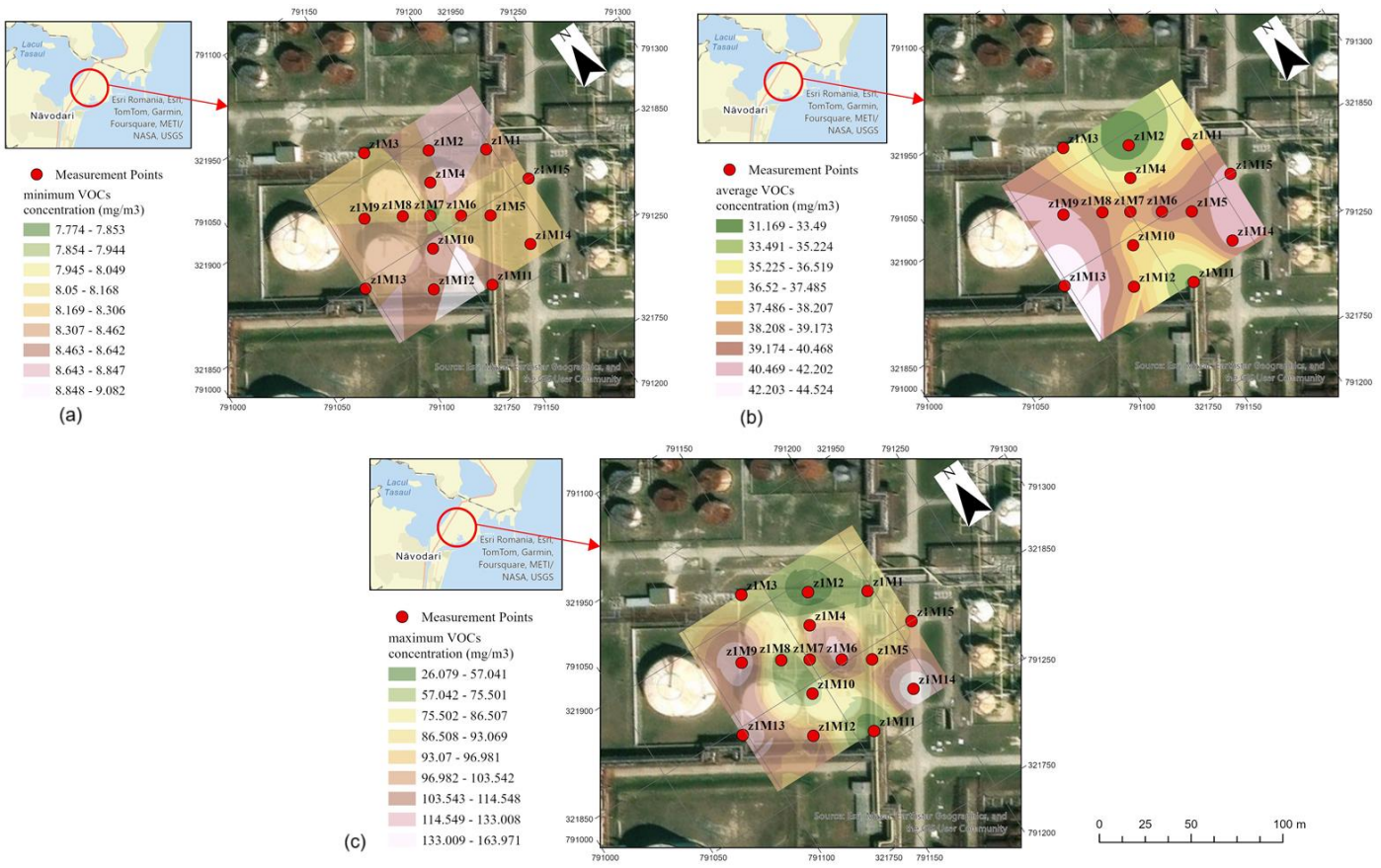

Figure 3. Maps reflecting the spatial distribution of the (**a**) minimum, (**b**) average, and (**c**) maximum concentrations in Zone 1.

Therefore, the maps provide information about the extent of pollution. Of course, more experiments must be done to adjust them.

II. The boxplots of all series are shown in Figure 4. They are built using the quartiles. The bottom (top) horizontal line corresponds to the first (third) quartile, Q_1 (Q_3), and the horizontal line inside the box is the median (Q_2). The whiskers represent the ranges for the data series' top and bottom 25% values, respectively, without considering the outliers.

Figure 5 (left) contains the correlogram and the PACF chart of the Z1M10 series. They show a first-order autocorrelation and partial autocorrelation of the Z1M10 series.

The examination of the correlograms of all series indicates that 7 series from Zones 1 and 2 and 5 from Zone 3 present a first-order autocorrelation. PACFs reveal significant partial autocorrelations (at a significance level of 95%) of the first order for 7 series from Zone 1 and 5 series from Zones 2 and 3.

The normality tests rejected the null hypothesis for 7 series in Zone 1, 8 in Zone 2, and 9 in Zone 3. Figure 6 displays the QQ plot of the Z1M14 series. The samples' quartiles (represented by dots) do not align with the theoretical (Gaussian) ones (represented by the line), showing that the data series is not normally distributed. The homoskedasticity The hypothesis was rejected for one, six, and four series from Zones 1, 2, and 3, respectively.

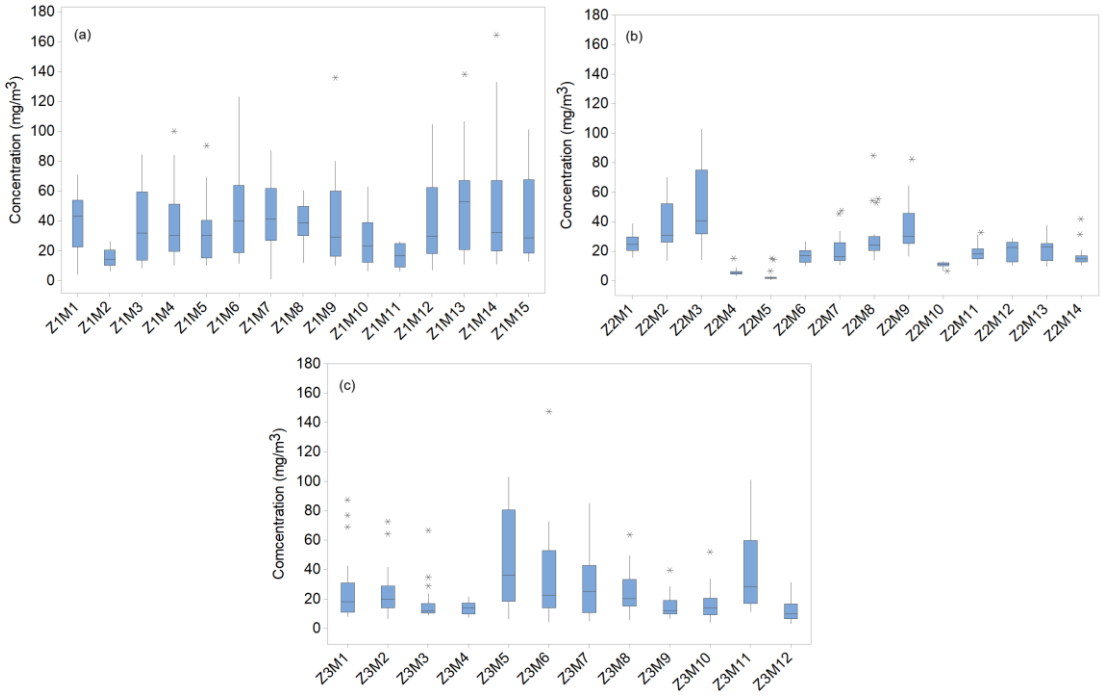

Figure 4. Boxplot of the series from (**a**) Zone 1, (**b**) Zone 2, (**c**) Zone 3. The stars represent the outliers.

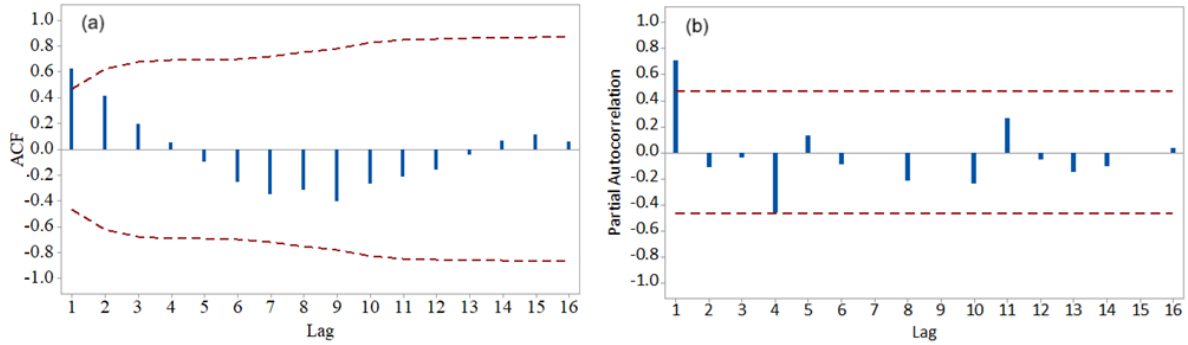

Figure 5. (**a**) ACF and (**b**) PACF of the Z1M10 series. The blue bars represent the ACF values at different lags. The dashed lines are the limits of the 95% confidence intervals.

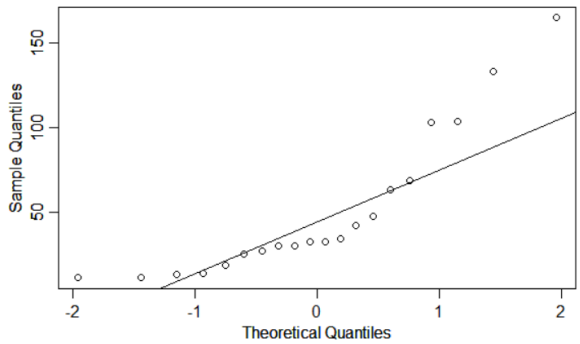

Figure 6. The QQ plot of the series Z1M10. The dots represent the sample's quantile. The line represents the theoretical (Gaussian) quantiles.

Table 2 summarizes the features from the above statistical analysis (performed according to point II from Methodology). It also contains the autocorrelation and partial autocorrelation orders and the number of outliers in the series recorded at each point. The rejection of the normality (homoskedasticity, respectively) hypothesis is marked by 0. These features will be utilized in the clustering stage.

Table 2. Features of the data series.

Zone 1	**M1**	**M2**	**M3**	**M4**	**M5**	**M6**	**M7**	**M8**	**M9**	**M10**	**M11**	**M12**	**M13**	**M14**	**M15**
Normality	1	1	1	1	0	0	1	1	0	1	0	0	1	0	0
Homoskedasticity	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1
ACF order	1	0	1	1	0	1	0	0	0	1	1	1	0	0	0
PACF order	1	0	1	1	0	1	0	0	0	1	1	1	0	0	0
Outlier	0	0	0	1	1	0	0	0	1	0	0	0	1	1	0
Zone 2	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	
Normality	1	1	1	0	0	1	0	0	0	0	1	0	1	0	
Homoskedasticity	0	0	1	1	1	1	0	0	0	0	1	1	1	1	
ACF order	0	0	0	0	0	0	1	1	1	1	0	1	1	1	
PACF order	0	0	0	0	0	0	1	1	1	1	0	0	0	1	
Outlier	0	0	0	1	3	0	2	3	1	1	1	0	0	2	
Zone 2	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12			
Normality	0	0	0	1	1	0	0	1	0	0	0	0			
Homoskedasticity	1	1	1	1	1	1	0	1	0	1	1	0			
ACF order	1	0	0	1	1	0	0	0	0	0	1	1			
PACF order	1	0	0	1	1	0	0	0	0	0	1	1			
Outlier	3	2	3	0	0	1	0	1	1	1	0	0			

III. The correlation matrix of the series in Zone 1 (2 and 3) presented in Figure 7 (Figure S4) indicates very few significant correlations.

	Z1M1	Z1M2	Z1M3	Z1M4	Z1M5	Z1M6	Z1M7	Z1M8	Z1M9	Z1M10	Z1M11	Z1M12	Z1M13	Z1M14
Z1M2	0.233													
Z1M3	0.433	0.531												
Z1M4	0.102	0.140	−0.181											
Z1M5	0.327	0.112	0.237	−0.134										
Z1M6	0.335	0.076	−0.094	0.034	0.375									
Z1M7	0.131	0.064	0.059	0.565	0.151	0.097								
Z1M8	0.103	0.480	0.379	0.444	0.253	0.379	0.550							
Z1M9	0.658	0.167	0.191	0.185	−0.285	−0.193	0.053	0.137						
Z1M10	0.067	0.297	0.214	0.349	−0.011	−0.195	0.232	0.289	−0.117					
Z1M11	0.477	0.317	0.231	0.342	−0.360	−0.536	0.138	−0.018	0.405	0.642				
Z1M12	0.290	0.009	0.100	0.237	0.012	0.038	0.166	0.421	0.072	−0.173	−0.351			
Z1M13	0.270	−0.064	−0.084	0.187	0.282	−0.234	0.200	0.020	0.066	0.276	0.249	0.219		
Z1M14	0.333	0.071	0.230	−0.179	0.053	−0.285	−0.043	−0.058	−0.330	0.158	0.023	−0.216	−0.117	
Z1M15	0.334	−0.060	0.079	−0.235	0.486	−0.025	0.305	0.157	−0.144	0.045	−0.021	−0.150	0.274	0.140

Figure 7. Correlation matrix of the series in Zone 1.

The distance correlation t-test of independence confirmed the results. The explanations for these low correlations could be the atmospheric calm during the period when the data series were collected and the relatively low height (1.4 m) at which the measurements were done. Moreover, the tanks are situated in an area surrounded by fences approximately 3 m high, which prevent significant air circulation.

The Anderson–Darling test returned *p*-values very close to zero, indicating that the samples (in a zone) are not extracted from the same population, so they do not follow the same repartition. The Mood's test rejected the null hypothesis, so one cannot accept the hypothesis that the series have the same median.

Figure 8 (Figure S5) displays the results of this test for Zone 1 (2 and 3).

The lowest dissimilarities are noticed between the series in Zone 1. Table 3 presents the *p*-values in the variance tests.

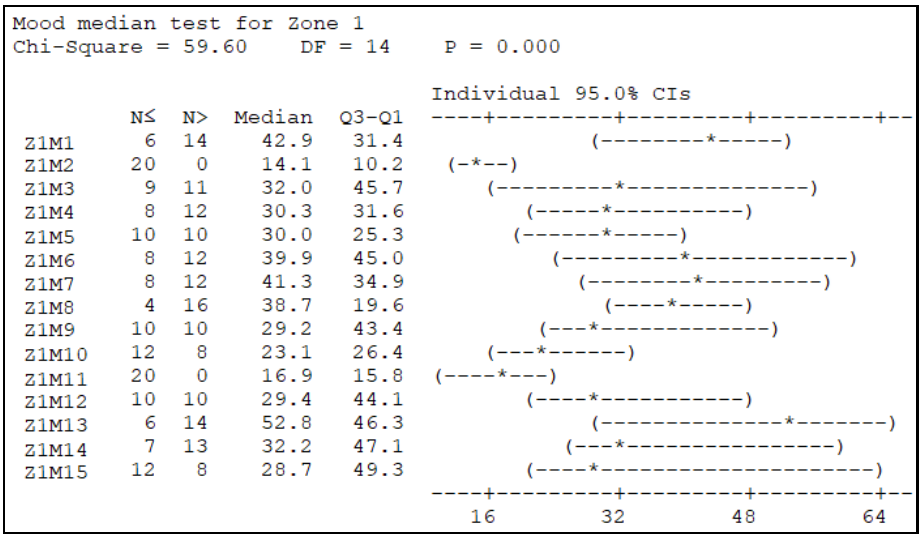

Figure 8. Results of the Mood median test for Zone 1.

They indicate that the hypothesis that all series in each zone have the same variance can be rejected, given that the *p*-values associated with both tests are very close to zero. This result indicates high outlier values, which lead to the variance increase.

Table 3. *p*-values and decision on the rejection of the variances' equality in the variance tests.

	Zone 1	**Zone 2**	**Zone 3**
Levene	3.059×10^{-4}	5.390×10^{-11}	4.219×10^{-16}
Fligner-Killeen	$<2.2 \times 10^{-16}$	1.934×10^{-11}	4.642×10^{-8}
Reject	yes	yes	yes

As a partial conclusion from part III, the series groups are inhomogeneous from the viewpoint of their properties as elements of each group.

IV. First, the sites in each zone were clustered, considering the raw data series.

To determine the number of clusters, we ran 23 algorithms implemented in the NbClust package in R. For Zone 1, 8 algorithms determined six as the optimum number of clusters, 6 determined 2 and 5, and the rest indicated other values for *k*. The results of the analysis of the internal clustering connectivity and clusters' stability are given in Table 4, rows 3–6.

Table 4. Highest internal connectivity and stability scores and the number of clusters.

	Internal Clustering Connectivity			**Clusters' Stability**		
Zone	Index	Score	*k*	Index	Score	*k*
1	Connectivity	3.1401	2	APN	0.0371	2
	Dunn	0.7462	2	ADM	0.2720	3
	Silhouette	0.2156	2	AD	3.1895	6
				FOM	0.9115	6
2	Connectivity	9.4143	2	APN	0.0000	2
	Dunn	0.6032	3	ADM	0.0000	3
	Silhouette	0.3583	2	AD	1.8039	6
				FOM	0.5100	6
3	Connectivity	2.9290	2	APN	0.0000	2
	Dunn	0.9335	4	ADM	0.0000	2
	Silhouette	0.3555	2	AD	1.9479	6
				FOM	0.6360	6

The best connectivity is obtained with 2 clusters, while, in terms of stability, $k = 6$ is the best choice according to AD, 3 according to ADM, and 2 according to APN. In the case of two clusters, the maximum ratio BSS/TSS is 17.75, indicating a low separation between clusters, and the AJ values are 0.867 and 0.878, showing their high stability. By comparison, selecting six clusters gave BSS/TSS = 64.4%, so there was a better separation between clusters.

The AJ values computed after 20 bootstrapping are between 0.600 and 0.780, indicating that two clusters are unstable and four stable. In conclusion, $k = 2$ is kept as the best choice. Moreover, the excessive fragmentation of the data series is avoided. Figure 9 shows the cluster plots for the series in Zone 1, for $k = 2$ and $k = 6$.

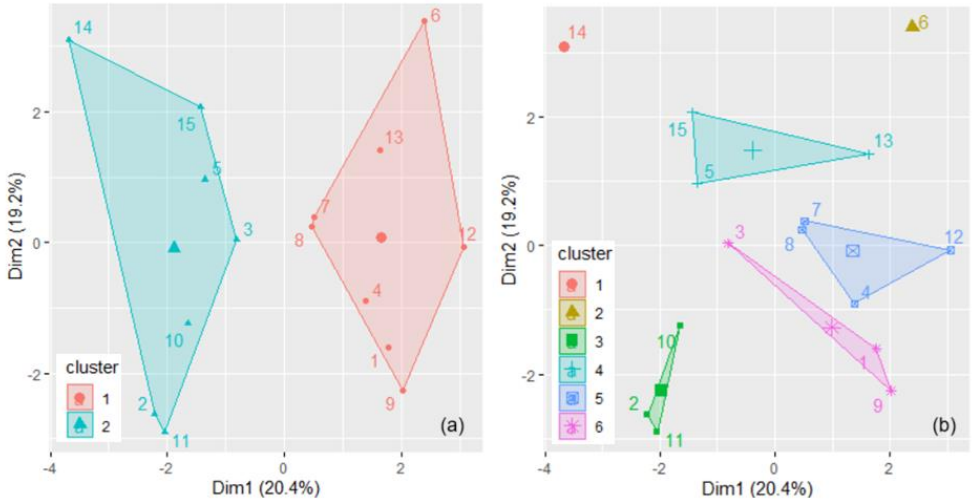

Figure 9. The cluster plots of the series in Zone 1 for (**a**) $k = 2$ and (**b**) $k = 6$.

For the series in Zone 2, the best number of clusters were 2 (10 algorithms), 3 (5 algorithms), and 6 (7 methods). The highest BSS/TSS was 82.7% for $k = 6$. The Connectivity and Silhouette indicate that the highest internal clustering connectivity is obtained for $k = 2$, whereas Dunn provided $k = 3$. The highest clusters' stability was determined for $k = 6$ by AD and FOM, $k = 2$ by APM, and $k = 3$ by ADM (Table 4—rows 7–10). The AJ values indicated one highly unstable cluster when $k = 6$. When $k = 2$ and $k = 3$, the clusters are stable. The cluster plots are presented in Figure 10.

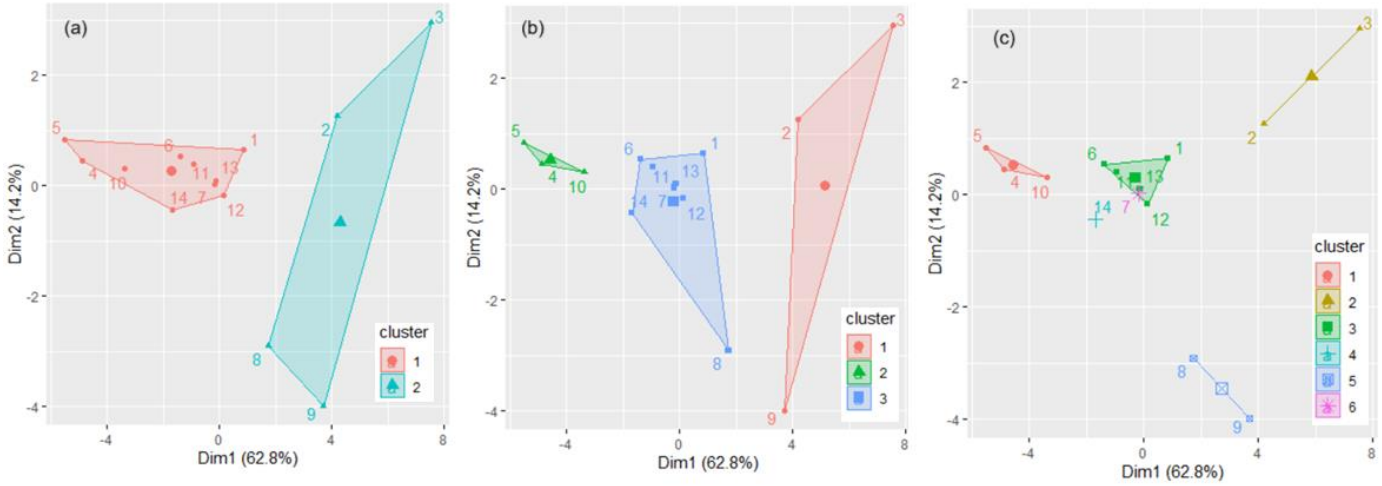

Figure 10. Cluster plots for the series in Zone 2 for (**a**) $k = 2$, (**b**) $k = 3$, and (**c**) $k = 6$.

For the series in Zone 3, the best number of clusters was found to be 1 (5) by one (one) algorithm, 2 by 8 procedures, 3 (and 6) by 5 (5) methods, and 4 by 3 methods. The rows 11–13 in Table 4 contain the optimal internal connectivity and stabilities (computed by varying the number of clusters between 2 and (6). It indicates the highest internal connectivity for $k = 2$. The highest clusters' stability was found when $k = 2$ or $k = 6$, but AJ indicated two unstable clusters when $k = 6$. Therefore, we selected $k = 2$ as the best number of clusters. The groups obtained when $k = 2$ and $k = 6$ are presented in Figure 11.

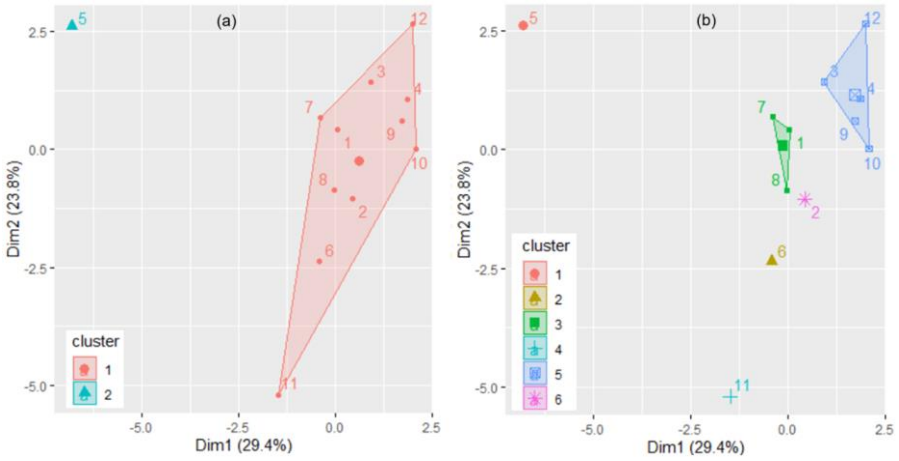

Figure 11. Cluster plots for the series in Zone 3 for (**a**) $k = 2$ and (**b**) $k = 6$.

The clustering in Figure 11a shows the highest variation in the series values from Z3M5, leading to a clear separation of this series from the others. Z3M5 presents a significantly high peak, which can be associated with the beginning of the revision period, followed by lower values of concentrations. The lowest variability is in the series Z3M11, situated at a significant distance from the other elements in the first cluster in Figure 11a. The values of other statistical characteristics—mean, standard deviation, normality, and homoscedasticity- balance the decision to cluster the data series in various groups.

In the second stage, the sites in each zone were clustered, considering the features' series determined in I and II. Specifically, the list of the features used is: F1-minimum, F2-maximum, F3—mean, F4—standard deviation, F5—Skewness, F6—kurtosis, F7—normality, F8—homoskedasticity, F9—ACF, F10—PACF, F11– outliers.

Running NbClust For the Features series in Zone 1, we saw that one algorithm found 1 as the best number of clusters, 10—2 clusters, 6—3 groups, 1—4 clusters, and 5—6 clusters, respectively. Based on the maximum principle, we selected $k = 2$. The validation measures (Table 5, rows 3–6, columns 5–7) indicated the highest stability for two clusters and various values for k for the highest internal connectivity (Table 5, lines 3–6, column 4). The AJ got the values 0.8716 for cluster 1 and 0.8749 for cluster 2, indicating very high stability of each cluster.

For the features series in Zone 2, the procedures implemented in NbClust returned the best number of clusters: 2 (3 algorithms), 3 (6 algorithms), 4 (4 algorithms), 5 (2 times), and 6 (7 times), respectively. The Dunn index, AD, and FOM indicate that $k = 6$ is the best choice from the viewpoint of stability and connectivity (Table 5, rows 7–10).

According to the AJ values, two clusters are unstable when $k = 6$. The internal measures Silhouette and ADM indicated $k = 5$ as the optimal number of clusters from the viewpoint of stability and connectivity, contradicting the results found running NbClust. Based on the majority principle and the above analysis, we kept $k = 6$ to balance the connectivity and stability criteria.

Table 5. Internal connectivity and stability of the clusters obtained using the Feature sets.

	Internal Clustering Connectivity			Clusters' Stability		
Zone	**Index**	**Score**	***k***	**Index**	**Score**	***k***
1	Connectivity	10.7210	2	APN	0.0000	2
	Dunn	0.8003	5	ADM	0.0000	2
	Silhouette	0.2882	3	AD	1.4794	6
				FOM	0.1832	2
2	Connectivity	6.2607	2	APN	0.0198	4
	Dunn	1.0088	6	ADM	0.1018	5
	Silhouette	0.3854	5	AD	0.9774	6
				FOM	0.4616	6
3	Connectivity	10.6345	2	APN	0.0248	2
	Dunn	0.8368	6	ADM	0.1513	6
	Silhouette	0.2385	2	AD	1.5051	2
				FOM	0.6955	6

For the Features series in Zone 3, one algorithm found $k = 1$ as the best number of clusters, 5 found $k = 2$, 7 found $k = 3$ and $k = 6$, and 2 found $k = 4$, respectively. Considering the stability and connectivity, most indices found $k = 2$ and $k = 6$ as the best choice (Table 5, rows 11–14). When $k = 6$, the AJ has the values 0.4833, 0.8000, 0.7250, 0.7000, 0.4000, and 0.9000, so the first and fifth clusters are unstable, the second, third, and fourth are stable, and the last one is highly stable. For $k = 3$, AJ values are between 0.6608 and 0.8068, indicating stable clusters. The clusters of the series of features in Zone 1 ($k = 2$), Zone 2 ($k = 6$), and Zone 3 ($k = 3$) are presented in Figure 12.

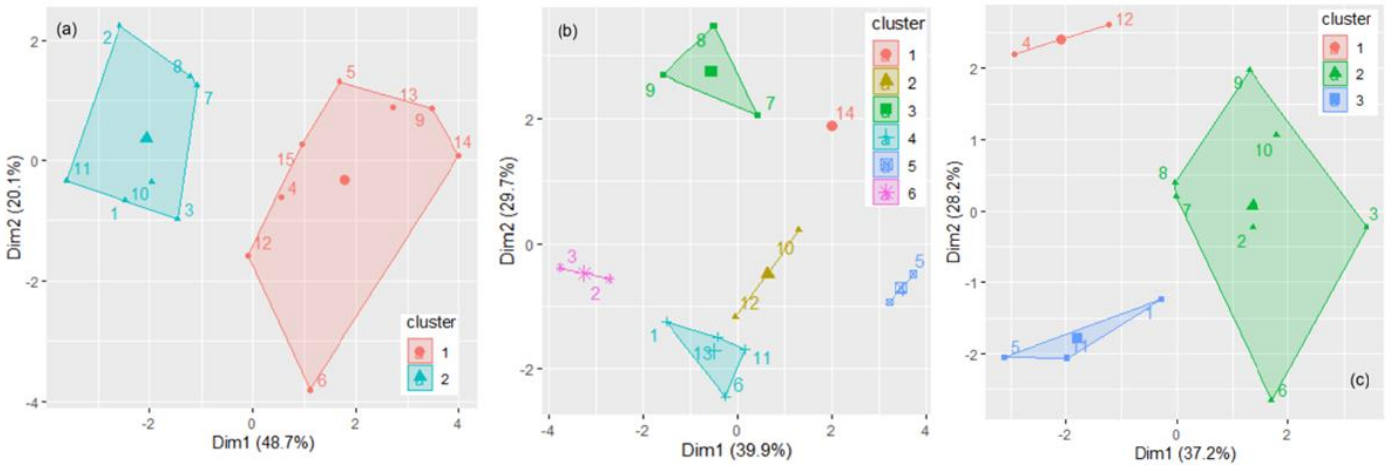

Figure 12. Clusters of the features' series in (**a**) Zone 1, (**b**) Zone 2, (**c**) Zone 3.

In Figure 12a, the first cluster contains the series recorded at Z1M1, Z1M2, Z1M3, Z1M7, Z1M8, Z1M10, and Z1M11, the rest belonging to the second cluster. Most points from the first cluster are situated on the border of Zone 1, and the other is on the perpendicular bisector of one of the sides. Therefore, the series of emissions from the tanks ''bordered'' by these points seems to have a similar behavior. The points from the second cluster are situated at a higher distance from the emission points. The AJ (0.876, 0.870) coefficients indicate a high stability of both clusters. Moreover, note that all of them are in the zone with the maximum concentrations of VOCs (Figure 3c).

The same remark applies to clustering the series in Zone 2 based on the features series after comparing the results with the charts in Figure S2c. The stations grouped in the same clusters in Figure 12c are characterized by low average concentrations (cluster 1) and

moderate to high maxima (cluster 2). By comparison, Figure 11 does not provide a clear understanding of the series classification.

V. The PCA was performed to determine the main features influencing the series classifications. First, the eigenvalues were computed together with the percentage of the variance explained by each and the percentage of the cumulative variance. Two methods were used to select the number of the principal components (PCs) to be retained: (a) keep the components whose eigenvalues are greater than one, and (b) keep the components until the cumulative percentage variance explained is high enough.

For Zone 1, we can keep either the first three PCs (the eigenvalues are greater than one) that explain 78.52% of the variance or the first four components, which explain 87.28%. For the second zone, we can keep the first three PCs (with the corresponding eigenvalues 4.3845, 3.2678, and 1.7975) that explain 85.908% of the variance or the first four, explaining 92.706 of it. Zone 3's first four components have eigenvalues greater than 1, explaining 90.4125% of the variance. The scree plots containing the percentage of explained variance by each PC are presented in Figure 13.

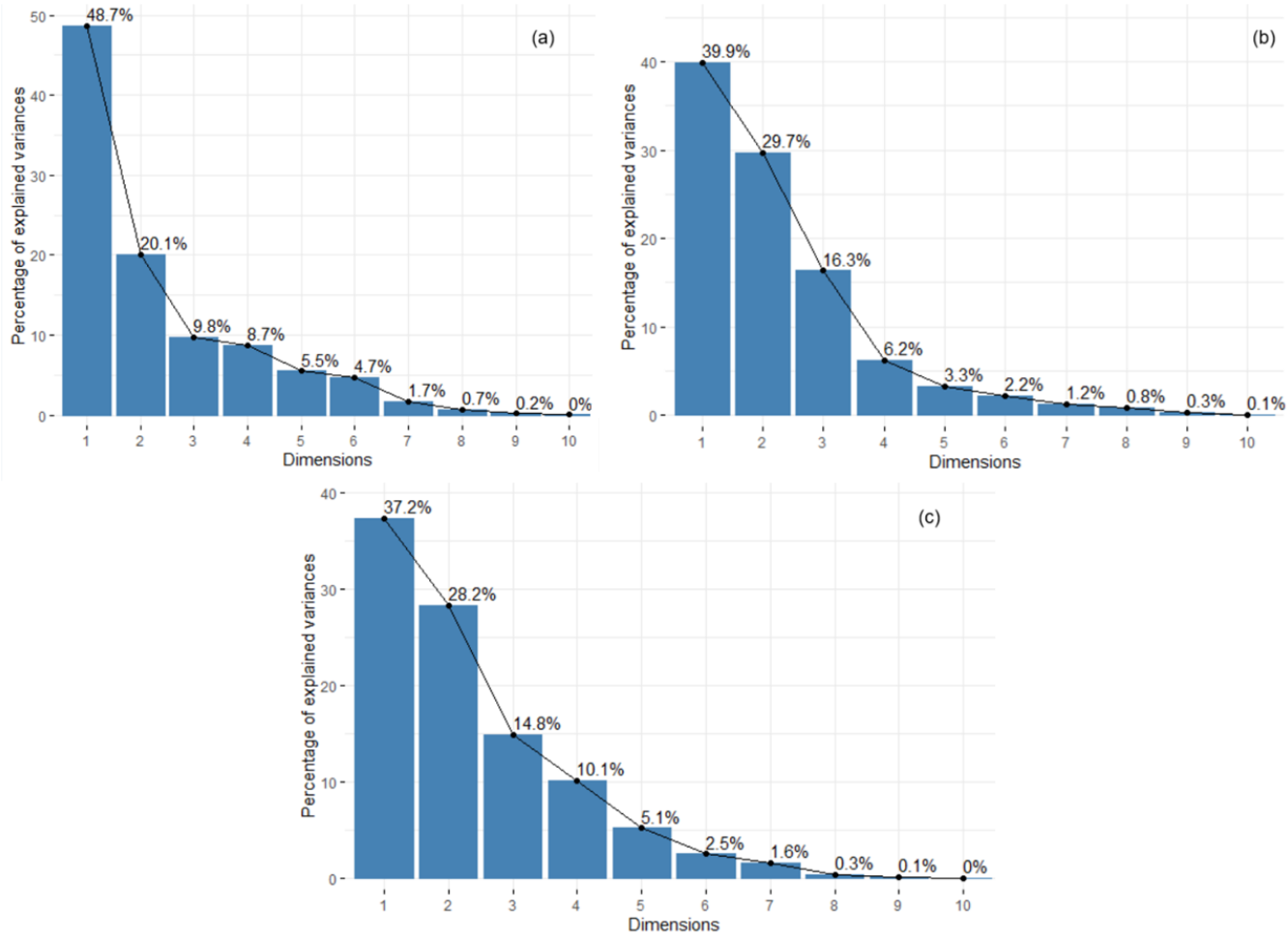

Figure 13. Screeplot for (**a**) Zone 1, (**b**) Zone 2, (**c**) Zone 3.

For Zone 1, Figure 14a shows the contributions of the variables on the first two dimensions in colors, with the lowest contributions in turquoise and the highest in dark orange. The contributions of the features on the third dimension are presented in Figure 14b. For example, the highest contribution on the first dimension is that of F2 (maximum), followed by F4 (standard deviation) and F6 (kurtosis).

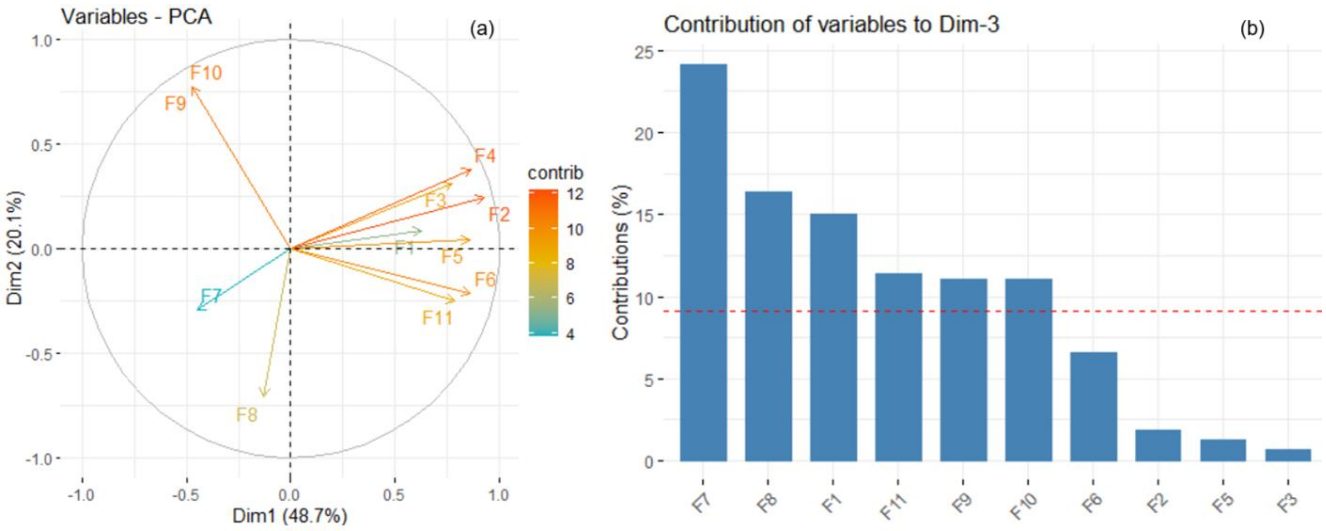

Figure 14. (**a**) Variables PCA, (**b**) Contributions of the variables to the third dimension.

The highest contributions on the second dimension are those of F9 (ACF), F10 (PACF), and F7 (normality), whereas on the third one—F7 (normality), F8 (homoskedasticity), and F1 (minimum). Summing up, it results that the main contributions on the first two dimensions are those of F2 (maximum), F4 (standard deviation), F9 (ACF), and F10 (PACF) (Figure 15a), whereas on all three dimensions, F9, F10, F2, and F4 (Figure 15b).

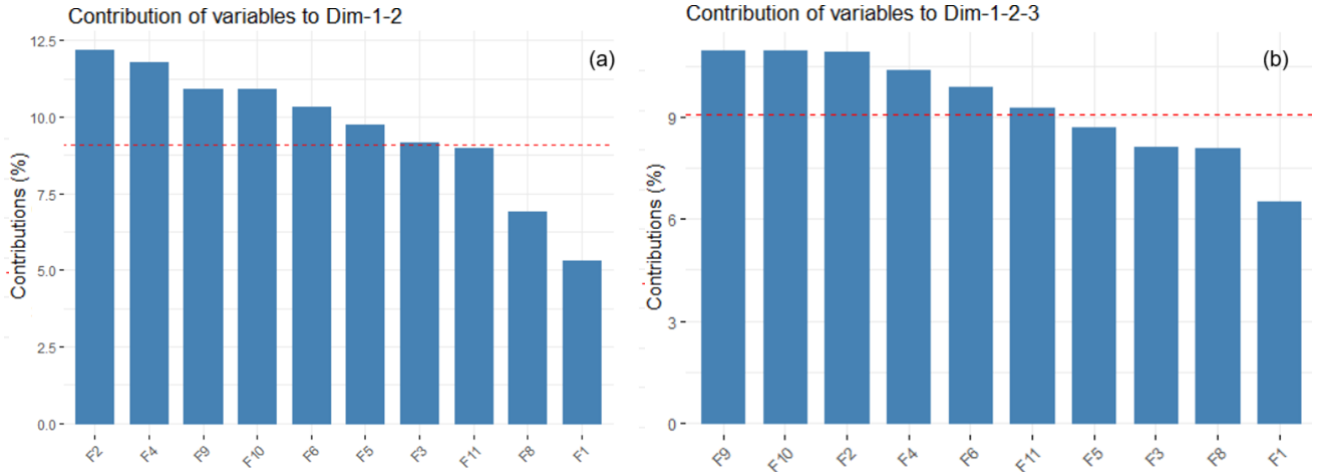

Figure 15. Contributions of the variables to (**a**) the first two and (**b**) the third dimension. The dashed line shows the expected average contribution. The most important contributions are those above the expected value.

The quality of representation for variables on the factor map is called cos2. It is displayed in Figure 16 for Zone, with different dimensions of the bullets and colors depending on the intensity. The highest contributions on dimension 1 are those of Sites 14, 9, 13 (on the positive direction of the first axis—belong to the first cluster in Figure 12a), and 11, 1, 10 (on the negative direction of the first axis—belong to the second cluster in Figure 12a.

Figures S6 and S7 present the detailed contributions of the variables (features) in Zones 2 and 3 on different dimensions. Figure 17 shows the contributions of each variable on the first 5 dimensions and the biplots. We summarize the findings below.

For Zone 2:

- The main contributions on Dim-1 are those of F3, F1, F2, and F4 ;

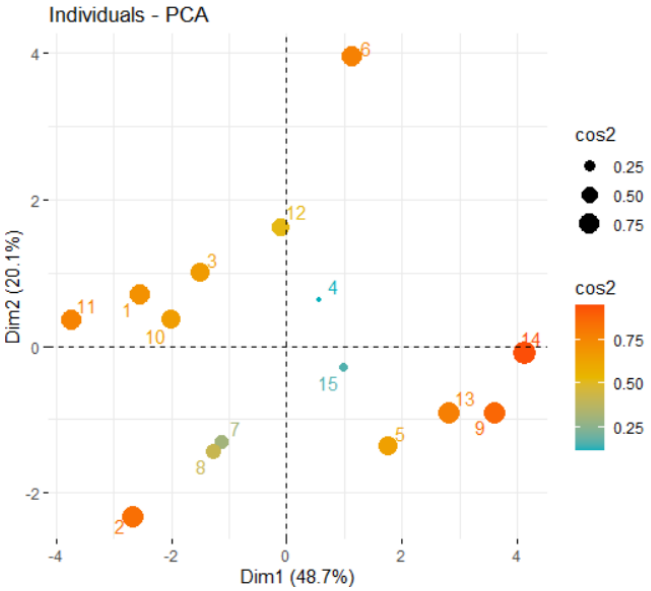

Figure 16. Individuals PCA for Zone 1.

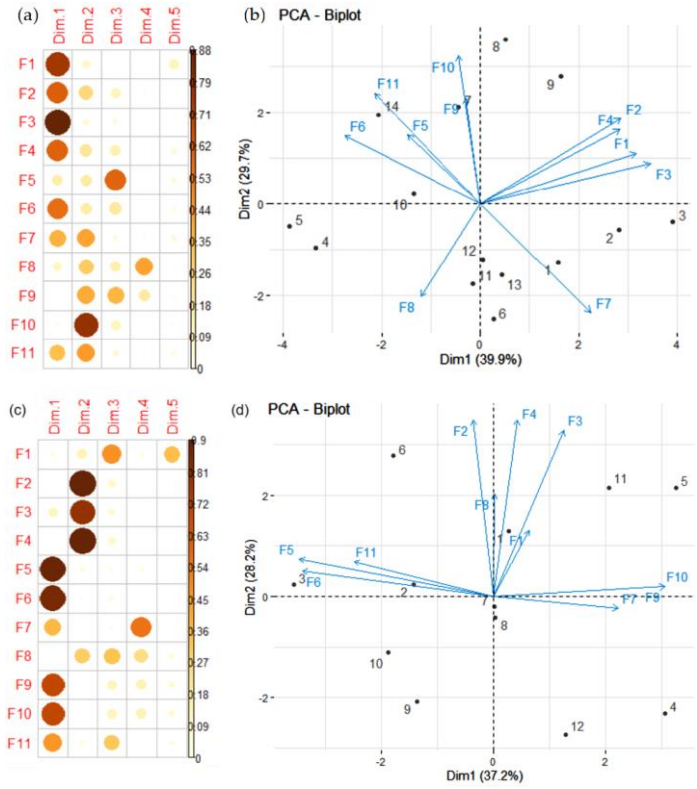

Figure 17. (**a**) Contributions of the features in Zone 2 on Dim1-Dim5, (**b**) PCA–biplot for Zone 2, (**c**) Contributions of the features in Zone 3 on Dim1-Dim5, (**d**) PCA–biplot for Zone 3.

- The main contributions on Dim-2 are those of F10, F12, F7, and F9 (i.e., PACF, outliers, normality, and ACF);
- The main contributions on Dim-1-2 are those of the average (F3), maximum (F2), minimum (F1), and PACF (F10);
- The most important contributions on Dim-1-2 are those of the average (F3), maximum (F2), standard deviation (F4), and skewness (F5);
- The most important contributions on Dim-1-2-3 are those of the average (F3), maximum (F2), standard deviation (F4), and skewness (F5);

- Considering the locations, the most important contributions on the dimensions are those of 3, 2 (on the positive direction of the first one), 5 and 4 (on the negative direction of the first one), 8 and 9 (on the positive direction of the second one), and 6 and 11 (on the negative direction of the second one).

For Zone 3:

- The main contributions on Dim-1 are those of F5, F6, F9, and F10 (i.e., skewness, kurtosis, ACF, and PACF);
- The most important contributions on Dim-2 are those of F2, F4, F3, and F8 (i.e., maximum, standard deviation, average, and homoskedasticity);
- The main contributions on Dim-1-2 are those of the skewness (F5), average (F3), maximum (F2), and standard deviation (F4);
- The essential contributions on Dim-1-2-3 are those of the standard deviation (F4), maximum (F2), average (F3), and skewness (F5).
- Considering the locations, the most important contributions to the dimensions are those of 5 and 4 (on the positive direction of the first one), 3 (on the negative direction of the first one), 6 and 11 (on the positive direction of the second one), and 12, 14, and 9 (on the negative direction of the second one).
- The above analysis points to the following ideas.

(1) The number of clusters must be rigorously selected and validated. The number of clusters must be chosen after comparing all criteria, given that stability and inner connectivity criteria can indicate as best k a value that does not concord with the number provided by the majority selection criterion (used by NbClust).

(2) The most significant contributions on the first three dimensions are maximum, standard deviation, and average (in this order). ACF and PACF also significantly contribute to the first or second dimensions. The comparison of the cluster obtained by the k-means algorithm on the raw and features series indicates a better classification provided by the last method, which considers the series' similar statistical characteristics. From a practical viewpoint, it points out unusually high emissions and the necessity of tank inspection, which can have significant implications for operational decision-making.

(3) A deeper investigation of the distributions of the series in each cluster might lead to determining similar underlying distributions. For example, based on the Anderson–Darling goodness-of-fit criterion, we found that the most suitable distribution for the series in the first cluster in Zone 1 is Wakeby (Table 6). In contrast, other distributions were fitted for the series in the second cluster. For the cluster with the highest number of elements in Zone 2, the best distributions fitted were Dagun and/or Wakeby.

Table 6. The Anderson–Darling test statistics (at the significance level of 5%). Cluster 1 in Zone 1.

Point	Z1M1	Z1M2	Z1M3	Z1M7	Z1M8	Z1M10	Z1M11
Statistics	0.20342	0.2991	0.3338	0.2969	0.1592	0.1532	0.7944

5. Discussions and Recommendations

The study's results underscore the complexity of pollutant dispersion in a tank farm.

As presented at the beginning of Section 4. Results, the statistical analysis and clustering results can be understood in correlation with field topography and the local conditions. The maps presented in the article, drawn taking into account the results from the three measurement campaigns and averaging them for each measurement point in each zone (as explained in Methodology), are informative for the pollution extent given the relatively low height at which the measurements were done, similar meteorological conditions, and the walls surrounding the tank area that do not permit a quick displacement of VOCs.

Spatial distribution maps offer real practical value because they indicate areas with localized emission on the study platform, which, although reduced in geographical size, includes zones with distinct operational characteristics for some tanks (some are in operation and others are under review). These differences in location are significant for the assessment of VOC emission patterns, especially during the revision period of the tanks where non-conformities have been identified (defects, lack of additional sealing elements), when emissions can spatially vary as a function of all the activities carried out in the area, as well as those carried out in the immediate vicinity.

The positions of the access roads create a corridor for pollutant dispersion. The statistical analysis indicates that VOC emissions predominantly propagated along these pathways, leading to pollutant accumulation at the ends of roads R4 and R9. Notably, the highest concentrations were observed at the points at the ends of the R4 access road, which connects to Zone 2. This finding suggests an interaction between emissions from the two zones and a possible dispersion of pollutants from both zones toward these critical points. The overlap of emissions from the two zones highlights the need for an integrated analysis of the entire access road network and reservoir configuration to understand the dispersion of fugitive emissions fully.

A classification based on statistical indicators can provide essential information about emissions variability and highlight areas that require immediate measures to reduce fugitive emissions. This approach has practical relevance, as it identifies statistically significant patterns that would otherwise have been difficult to observe if only primary data (momentary concentrations) had been used. Moreover, to improve the spatial distribution mapping and better contextualize VOC's distribution patterns, future studies must consider hourly records (that capture the short-term fluctuations of the emissions).

Since the emissions are directly attributable to specific tank maintenance activities, by narrowing the study objectives to this specific operational phase, the results can provide applicable insights for improving local emission control strategies and monitoring protocols in similar situations.

Based on the clustering results, the maintenance schedules or tank designs might be adapted for reducing the pollution. Such measures involve:

- *Adaptation of tank design*: The classification based on statistical characteristics highlights aspects that may influence emissions, such as the volatility of the stored products or the design of safety elements of the equipment. Refineries can use this information to modernize tanks by implementing more efficient characteristic elements or by installing advanced vapour recovery systems in high-emission areas.
- *Control measure planning*: Areas with high VOC concentrations may be subject to stricter monitoring and specific interventions, such as using protective barriers, increasing local ventilation, or installing equipment to capture the vapors.
- *Data-driven decision making* in situ: The refineries can integrate cluster analysis into their monitoring and reporting systems. This integration can support operational and strategic decisions, providing a clear perspective on critical areas and periods concerning emissions.

The main directions through which our results can support the strategies to be applied in the maintenance scheduling of the tank of the equipment in the reservoir parks and the compliance with the environmental regulations are:

- *Optimizing the equipment monitoring programs*. The statistical characteristics of the series and the series clustering indicate areas with high levels of emissions or pronounced variability in VOC concentrations. This information can be used by companies to prioritize verification activities in critical areas, thus reducing the impact of emissions on the environment. For example, for areas 1 and 2, where maximum

values and standard deviation were high, operations carried out to establish the operational safety level of the industrial equipment (which are periodically performed to verify the integrity and/or tightness of valves, pipes, and, implicitly, tanks, as well as hazard warning systems) must be prioritized, under stricter safety measures, or scheduled during periods with favourable meteorological conditions.

- *Taking measures for reducing emissions as much as possible*. By classifying the distribution of VOC emissions, the areas that require additional mitigation strategies are highlighted. For example, points/places with maximum concentrations or high variability can benefit from improved sealing methods by applying best available techniques (BAT) based on the requirements of Directive 2010/75/EU implemented by European Commission decision 2014/738/EU and the use of vapor recovery systems or more stringent operational controls during verification periods [48].
- *Removing non-conformities based on environmental regulations*. Digital representations of pollutant distribution (spatial distribution maps) and clustering provide a solid framework for observing compliance with air quality standards. By quantifying accepted emission levels in the industrial environment (the accepted average hourly concentration for the volatile non-methane organic compounds is between 0.15 and 10 g/Nm^3 during the operation period) and identifying critical areas, refineries can develop reporting mechanisms and develop action plans to comply with the legal requirements.
- *Continuous monitoring*. The methodologies used in this study (statistical feature analysis and clustering) can serve in developing real-time monitoring systems that can be deployed during the revision periods to provide early warnings and support response actions to increases in emissions, thereby improving operational safety and decreasing environmental pollution.

International studies on refineries indicate a wide range of VOC concentrations, influenced by factors such as the type of stored products, control measures implemented, and local climatic conditions. For example, the VOC emission inventory by Lv et al. [49] at the Shandong refinery in China (with an annual crude oil processing capacity of 12 million tons) found that petroleum product storage areas were the largest sources of VOC emissions, accounting for 56.4% of the total emissions at the refinery, followed by petroleum loading/unloading operations, the wastewater collection and treatment system, the ventilation outlets of the installations, and leaks from industrial equipment. Another study conducted by Rovira et al. [50] assessed the environmental impact of the chemical/petrochemical complex, which includes the crude oil refining sector in Tarragona from 2015 to 2019, as well as the risks arising near these industrial zones.

Our findings are in concordance with the findings of Bodor et al. [51], Sanda et al. [52], and Barbes et al. [53], reporting high pollution produced by the oil refining complex Ploiesti and Năvodari, Romania. No matter the authors or regions for which the research was performed, it was proved that pollution provoked by refineries is a general issue. Therefore, the governmental organisms must impose measures to be implemented and strictly followed in any crude oil refining complex. Nevertheless, identifying areas with maximum values allows for the adoption of tailored measures to reduce pollution, which represents an important step toward compliance with international/national/regional environmental standards. In this context, it is necessary for industrial installations and equipment to be regularly inspected and repaired to ensure proper functioning.

The results of our study provide a valuable tool for decision-makers by identifying critical areas with high levels of VOC emissions. This information can be used to develop much more effective local environmental policies, such as:

- Setting stricter requirements for emission monitoring during industrial equipment inspection periods.

- Creating best practice guidelines for reducing emissions at the analyzed refinery.
- Identifying major VOC emission sources through the research study, which can assist the economic operator in implementing specific measures in direct correlation with compliance with national and international regulatory standards.

Thus, we propose:

Pollutants Monitoring and Management:

- Develop better monitoring systems to assess industrial pollutant emissions based on refinery equipment, layout, and operating conditions, enabling more effective management.
- Install automated meters and computer systems on process units for continuous monitoring of losses and ambient air quality, enabling rapid response to pollution.
- Use VOC vapor recovery systems at the beginning of the maintenance period to transfer existing vapors to other equipment or retention spaces.
- Perform rigorous inspections of the floating roofs inside the storage tanks.
- Using high-quality gaskets and seals at the equipment access points is crucial to prevent VOC leaks into the atmosphere. In some instances, inert atmospheres inside the tanks during maintenance prevent oxidation and the release of VOCs into the terrestrial atmosphere.

Operational planning:

- Performing inspection/maintenance work during periods with favorable weather conditions for controlled dispersion;
- Prioritizing investments in emission reduction technologies for these areas;
- Optimizing tank design to minimize emissions**.** For example, use larger tanks for crude oil and refined products, fitted with floating roofs and pontoons to reduce emissions.
- Installing vapor recovery systems in areas with high emissions.

Other solutions for pollution reduction in the refinery areas include [54–59]:

- Larger Refining Units and Integrated Processes: Implementing larger refining units and combining the processes that will reduce crude oil/product losses, process fuel consumption, and cooling water, leading to lower air pollution.
- Improve the Crude Oil Pre-Treatment by reducing salt content through a unified pretreatment system, both in the oil fields and the refinery, to minimize contaminated effluent water produced in the technological process. This method will reduce VOC emissions and wastewater discharge.
- Desulfurization of Fuels: Continue improving desulfurization technologies for engine and boiler fuels, especially for heavy residual stocks, and integrate sulfur recovery units to reduce SO2 and other sulfur compounds in emissions.

Scientists should create a pollution index to track ecosystem pollution from refineries, with dedicated systems to detect and mitigate environmental harm.

These proposed management strategies for mitigating air pollution will offer a significant drop in the whole air pollution.

The study highlights the need for proactive strategies to address VOC industrial emissions more responsibly. This is essential not only for compliance but also for improving public perception of industrial activities, thus contributing to sustainable development.

Further research directions include:

1. Extend the comparative spatio-temporal analysis for VOC emissions/immissions by conducting measurements on a daily basis over periods that cover both warm and cold seasons to understand the influence of seasonal variations on VOC emissions.

2. Extend monitoring at the same locations (the oil refining complex) by including other atmospheric pollutants, such as particulate matter (PM), nitrogen oxides (NO_x), and sulfur dioxide (SO_2), which we presented in our previous studies from monitoring campaigns conducted in urban areas and coastal zones near industrial platforms [25,26,28,60,61].
3. Develop and validate predictive models based on artificial intelligence, such as neural networks or machine learning methods, to anticipate VOC emissions based on operational activities and meteorological conditions.
4. Integrate in situ data with satellite data to obtain the best maps showing the pollution extent.
5. Provide digital representation/mapping of emissions over a broader geographical area, including residential areas adjacent to refineries, to assess the impact on local communities.
6. Analyze the long-term impact on VOC emissions and other pollutants on the people's health in the neighboring towns and villages.
7. Evaluating the costs associated with emission reductions versus the benefits for public health and the environment, integrating data on cumulative exposure to multiple pollutants.

6. Conclusions

Analyzing pollution dissipation is a complex process due to anthropic activity and natural landscapes. The research conducted using statistical methods provides a good framework for understanding the behavior of VOC emissions in a petroleum storage park. It focuses on the maintenance period of the storage tanks—a critical time when VOC emissions are likely at their peak. Maintenance activities in refineries often lead to significant increases in VOC emissions, which can contribute to air pollution and pose risks to the health of workers and nearby communities.

The results, derived from the records from 41 sampling points during three maintenance periods of 14 storage tanks with capacities ranging from 1000 to 10,000 m^3, not only contribute to understanding the pollution effect produced by the distribution of VOC emissions in these specific areas but also to the development of emission control strategies by improving industrial practices and reducing the environmental and human health impact within the refinery.

The study took into account only the VOC concentrations. The uneven pollutants' emission was emphasized by statistical tests of the sets of series and confirmed by the classification provided by the k-means algorithm. It was shown that the data series presents a high variability, and grouping them into clusters based on their statistical properties is a valuable alternative to using the raw series, since the former approach prioritizes the maxima, putting in evidence the necessity of pollution reduction if the admissible limits are surpassed.

The classification based on statistical indicators can provide essential information about emissions variability and highlight areas that require immediate measures to reduce fugitive emissions. This approach has practical relevance, as it identifies statistically significant patterns that would otherwise have been difficult to observe if only primary data (momentary concentrations) had been used.

Our approach emphasizes that the most important statistical features that contribute to the first three dimensions obtained after performing the PCA are maximum, standard deviation, and average. To validate our approach, other research must be performed using different algorithms and various types of series. The study will also be extended by taking into account the meteorological factors.

Although this study is focused only on VOC emissions, the methodology used—especially statistical analysis and clustering—can be extended to other pollutants. Future research should address these limitations by integrating meteorological and seasonal factors, extending the temporal and spatial scope, and incorporating other pollutants for an extended analysis. For example, integrating data on particulate matter (PM), NO_x or SO_2 could provide a more comprehensive understanding of the interactions between pollutants and their cumulative impact. Predictive models will be developed to incorporate other pollutants and their interactions.

Variation in the pollutants' concentration in vertical is another direction that deserves to be investigated and will be addressed in future studies based on a wider database collected hourly.

The research provides solutions to be implemented for improving air quality in industrial environments and reducing the impact on the health of workers as well as the population living near industrial plants. We underline the practical relevance of the study's results for operational decision-making in the analyzed refinery and to support the development of more effective strategies for air quality management in similar industrial environments.

Supplementary Materials: The following supporting information can be downloaded at: www.mdpi.com/article/10.3390/app142411921/s1, Figure S1: Basic statistics of the series recorded at the locations from Zones (a) 1, (b) 2, and (c); Figure S2: Maps reflecting the spatial distribution of the minimum, average, and maximum concentrations in Zone 2; Figure S3: Maps reflecting the spatial distribution of the minimum, average, and maximum concentrations in Zone 3; Figure S4: Correlation matrix for the series in (a) Zone 2 and (b) Zone 3; Figure S5: Mood's median test for (a) Zone 2, (b) Zone 3; Figure S6: Contribution of variables in Zone 2 to (a) Dim-1 (b) Dim-2, (c) Dim-3, (d) Dim-1-2, (e) Dim-1-2, (f) Dim-1-2-3; Figure S7: Contribution of variables in Zone 3 to (a) the Dim-1 (b) Dim-2, (c) Dim-3, (d) Dim-1-2, (e) Dim-1-2, (f) Dim-1-2-3; Table S1: Zone 1: Basic statistics; Table S2: Zone 2: Basic statistics; Table S3: Zone 3: Basic statistics.

Author Contributions: Conceptualization, S.-B.B. and C.Ș.D.; methodology, A.B. and C.Ș.D.; software, A.B. and C.Ș.D.; validation, A.B. and L.B.; formal analysis, S.-B.B. and L.B.; investigation, A.B. and C.Ș.D.; resources, S.-B.B. and L.B.; data curation, A.B.; writing—original draft preparation, A.B. and L.B.; writing—review and editing, A.B. and L.B.; visualization, L.B. and C.Ș.D.; supervision, A.B.; project administration, A.B.; funding acquisition, S.-B.B., C.Ș.D., A.B. and L.B. All authors have read and agreed to the published version of the manuscript.

Funding: The work was supported by the Technical University of Civil Engineering of Bucharest for the first two authors.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

Conflicts of Interest: Data will be available on request from the authors.

References

1. Evans, J.; Levy, J.; Hammitt, J.; Santos-Burgoa, C.; Castillejos, M.; Ramírez, M.C., Ávila, M.H., Rodríguez, H.R., Bracho, L.R.; Trespalacios, P.S.; et al. Health benefits of air pollution control. In *Air Quality in the Mexico Megacity: An Integrated Assessment*; Molina, L.T., Molina, M.J., Eds.; Springer: Dordrecht, The Netherlands, 2002.
2. Collins, J.J.; Ireland, B.; Buckley, C.F.; Shepperly, D. Lymphohaematopoeitic cancer mortality among workers with b2`1enzene exposure. *Occup. Environ. Med*. **2003**, *60*, 676–679.

3. Falzone, L.; Marconi, A.; Loreto, C.; Franco, S.; Demetrios, A.; Libra, M. Occupational exposure to carcinogens: Benzene, pesticides and fibers (Review). *Molec. Med. Rep*. **2016**, *14*, 4467–4474.
4. Snyder, R. Leukemia and benzene. *Int. J. Environ. Res. Public Health* **2012**, *9*, 2875–2893.
5. David, E.; Niculescu, V.-C. Volatile Organic Compounds (VOCs) as Environmental Pollutants: Occurrence and Mitigation Using Nanomaterials. *Int. J. Environ. Res. Public Health* **2021**, *18*, 13147.
6. Chen, C.H.; Chuang, Y.C.; Hsieh, C.C.; Lee, C.S. VOC characteristics and source apportionment at a PAMS site near an industrial complex in central Taiwan. Atmos. *Pollut. Res*. **2019**, *10*, 1060–1074.
7. Shusterman, D. The effects of air pollutants and irritants on the upper airway. *Proc. Am. Thorac. Soc*. **2011**, *8*, 101–105.
8. Bărbulescu, A.; Barbeş, L. Models for the pollutants correlation in the Romanian Littoral. *Rom. Rep. Phys.* **2014**, *66*, 1189–1199.
9. Milazzo, M.F.; Ancione, G.; Lisi, R. Emissions of volatile organic compounds during the ship-loading of petroleum products: Dispersion modelling and environmental concerns. *J. Environ. Manag*. **2017**, *204*, 637–650.
10. Saikomol, S.; Thepanondh, S.; Laowagul, W.; Malakan, W.; Keawboonchu, J.; Kultan, V. Characteristics and dispersion modeling of VOCs emission released from the tank farm of petroleum refinery complex. *Environ. Asia* **2021**, *14*, 1–12.
11. Karbasi, A.; Khoramnezhadian, S.; Asemi Zavareh, S.A.; Pejman Sani, G. Determination of the emission rate and modeling of benzene dispersion due to surface evaporation from an oil pit. *J. Air Pollut. Health* **2018**, *3*, 155–166.
12. Shahbazi, H.; Taghvaee, S.; Hosseini, V.; Afshin, H. A GIS based emission inventory development for Tehran. *Urban Clim.* **2016**, *17*, 216–229.
13. Shie, R.H.; Chan, C.C. Tracking hazardous air pollutants from a refinery fire by applying on-line and off-line air monitoring and back trajectory modeling. *J. Hazard. Mater.* **2013**, *261*, 72–82.
14. Hong, N.; Liu, A.; Zhu, P.; Zhao, X.; Guan, Y.; Yang, M.; Wang, H. Modelling benzene series pollutants (BTEX) build-up loads on urban roads and their human health risks: Implications for stormwater reuse safety. *Ecotox. Environ. Safe*. **2018**, *164*, 234–242.
15. Liu, J.; Li, W.Q. A Long-Term Modelling Study of Ventilation and VOC Distribution in Multi-family Residential Buildings in the Severe Cold Region of China. *Int. J. Vent.* **2011**, *10*, 217–226.
16. Feustel, H.E. COMIS—An international multizone air-flow and contaminant transport model. *Energ. Build.* **1999**, *30*, 3–18.
17. Zapata-Marin, S.; Schmidt, A.M.; Ho, V.; Labrèche, F.; Lavigne, E.; Parent, M.-É.; Goldberg, M.S. Spatial modeling of ambient concentrations of volatile organic compounds in Montreal, Canada. *Env. Epidemiol.* **2022**, *6*, e226.
18. Jung, S.W.; Lee, K.; Cho, Y.S.; Choi, J.H.; Yang, W.; Kang, T.S.; Park, C.; Kim, G.B.; Yu, S.D.; Son, B.S. Association by Spatial Interpolation between Ozone Levels and Lung Function of Residents at an Industrial Complex in South Korea. *Int. J. Environ. Res. Public Health.* **2016**, *13*, 728.
19. Li, S.; Xi, W.; Li, C.; Bi, T. Study on pollutant model construction and threedimensional spatial interpolation in soil environmental survey. *Conf. Ser. Earth Environ. Sci*. **2020**, *467*, 012154.
20. Chiritescu, R.-V.; Luca, E.; Iorga, G. Observational study of major air pollutants over urban Romania in 2020 in comparison with 2019. *Rom. Rep. Phys*. **2024**, *76*, 702.
21. Rosianu, A.-M.; Leru, P.M.; Stefan, S.; Iorga, G.; Marmureanu, L. Six-year monitoring of atmospheric pollen and major air pollutant concentrations in relation with meteorological factors in Bucharest, Romania. *Rom. Rep. Phys.* **2022**, *74*, 703.
22. Dumitru, A.; Olaru, E.-A.; Dumitru, M.; Iorga, G. Assessment of Air Pollution by Aerosols over a Coal Open-Mine Influenced Region in Southwestern Romania. *Rom. J. Phys.* **2024**, *69*, 801.
23. Antonescu, N.N.; Stănescu, D.-P.; Calotă, R. CO_2 Emissions Reduction through Increasing H_2 Participation in Gaseous Combustible—Condensing Boilers Functional Response. *Appl. Sci.* **2022**, *12*, 3831.
24. Bodor, Z.; Bodor, K.; Keresztesi, Á.; Szep, R. Major air pollutants seasonal variation analysis and long-range transport of PM10 in an urban environment with specific climate condition in Transylvania (Romania). *Environ. Sci. Pollut. Res.* **2020**, *27*, 38181–38199.
25. Bărbulescu, A.; Barbeș, L. Mathematical modeling of sulfur dioxide concentration in the Western part of Romania. *J. Environ. Manag.* **2017**, *204*, 825–830.
26. Bărbulescu, A.; Barbeș, L. Modeling the carbon monoxide dissipation in Timisoara, Romania. *J. Environ. Manag.* **2017**, *204*, 831–838.
27. Bărbulescu, A.; Saliba, Y. Sensitivity Analysis of the Inverse Distance Weighting and Bicubic Spline Smoothing Models for MERRA-2 Reanalysis PM2.5 series in the Persian Gulf Region. *Atmosphere* **2024**, *15*, 748.
28. Bărbulescu, A.; Barbeş, L. Statistical Assessment and Modeling of Benzene Level in Atmosphere in Timiş County, Romania. *Int. J. Environ. Sci. Technol*. **2022**, *19*, 817–828.

29. Council Decision (EU) 2017/1757 of 17 July 2017 on the Acceptance on Behalf of the European Union of an Amendment to the 1999 Protocol to the 1979 Convention on Long-Range Transboundary Air Pollution to Abate Acidification, Eutrophication and Ground-Level Ozone. Available online: https://eur-lex.europa.eu/eli/dec/2017/1757/oj (accessed on 30 November 2024).
30. Law no. 264 of December 20, 2017 Regarding the Establishment of Technical Requirements for the Limitation of Emissions of Volatile Organic Compounds (VOC) Resulting from the Storage of Gasoline and Its Distribution from Terminals to Gasoline Distribution Stations, as Well as During Fueling of Motor Vehicles at Gasoline Stations. Available online: https://lege5.ro/gratuit/gi3dinjvg4yq/legea-nr-264-2017-privind-stabilirea-cerintelor-tehnice-pentru-limitarea-emisiilor-de-compusi-organici-volatili-cov-rezultati-din-depozitarea-benzinei-si-din-distributia-acesteia-de-la-terminale-la-st#:~:text=Legea%20nr.-,264%2F2017%20privind%20stabilirea%20cerin%C5%A3elor%20tehnice%20pentru%20limitarea%20emisiilor%20de,autovehiculelor%20la%20sta%C5%A3iile%20de%20benzin%C4%83 (accessed on 4 December 2024). (In Romanian)
31. Shapiro, S.S.; Wilk, M.B. An analysis of variance test for normality (complete samples). *Biometrika* **1965**, *52*, 591–611.
32. Field, A. *Discovering Statistics Using SPSS*; SAGE Publications: Los Angeles, CA, USA, 2009.
33. Levene, H. Robust tests for equality of variances. In *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*; Olkin, I., Ed.; Stanford University Press: Stanford, CA, USA, 1960; pp. 278–292.
34. Székely, G.J.; Rizzo, M.L. The distance correlation t-test of independence in high dimension. *J. Multivar. Anal.* **2013**, *117*, 193–213. https://doi.org/10.1016/j.jmva.2013.02.012.
35. kSamples-Package: The Package kSamples Contains Several Nonparametric K-Sample Tests and Their Combinations over Blocks. Available online: https://www.rdocumentation.org/packages/kSamples/versions/1.2-10/topics/kSamples-package (accessed 28 August 2024).
36. Brown, G.W.; Mood, A.M. Homogeneity of several samples. *Amer. Statist.* **1948**, *2*, 22.
37. Conover, W.J.; Johnson, M.E.; Johnson, M.M. A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data. *Technometrics* **1981**, *23*, 351–361.
38. K-Mean: Getting the Optimal Number of Clusters. Available online: https://www.analyticsvidhya.com/blog/2021/05/k-mean-getting-the-optimal-number-of-clusters/ (accessed 16 June 2023).
39. Hartigan, J.A.; Wong, M.A. Algorithm AS 136: A k-means clustering algorithm. *J. R. Stat. Soc. C Appl. Stat.* **1979**, *28*, 100–108.
40. Charrad, M.; Ghazzali, N.; Boiteau, V.; Niknafs, A. Determining the Best Number of Clusters in a Data Set. Available online: https://cran.r-project.org/web/packages/NbClust/NbClust.pdf (accessed 9 October 2024).
41. Brock, G.; Pihur, V.; Datta, S.; Datta, S. clValid: An R Package for Cluster Validation. *J. Stat. Softw.* **2008**, *25*, 1–22. https://www.jstatsoft.org/article/view/v025i04
42. Datta, S.; Datta, S. Comparisons and validation of statistical clustering techniques for microarray gene expression data. *Bioinformatics* **2003**, *19*, 459–466.
43. Soetewey, A. Stats and R. The Complete Guide to Clustering Analysis: K-Means and Hierarchical Clustering by Hand and in R. Available online: https://statsandr.com/blog/clustering-analysis-k-means-and-hierarchical-clustering-by-hand-and-in-r/ (accessed on 20 June 2023).
44. Yousuf, F. Difference Between K-Means and Hierarchical Clustering. Available online: https://medium.com/@wazirip hareeyda/difference-between-k-means-and-hierarchical-clustering-edfec55a34f8 (accessed on 2 October 2024).
45. Weber, C.M.; Ray, D.; Valverde, A.A.; Clark, J.A.; Sharma, K.S. Gaussian mixture model clustering algorithms for the analysis of high-precision mass measurements. *Nucl. Instrum. Meth. Phys. Res. A* **2022**, *1027*, 166299.
46. Yadav, A.K. Means Clustering vs Gaussian Mixture. Available online: https://medium.com/@amit25173/k-means-clustering-vs-gaussian-mixture-bec129fbe844 (accessed on 2 December 2024).
47. Jolliffe, I.T. *Principal Component Analysis*, 2nd ed.; Springer-Verlag: New York, NY, USA, 2002.
48. EUR-Lex. Available online: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32014D0738 (accessed on 2 December 2024).
49. Lv, D.; Lu, S.; Tan, X.; Shao, M.; Xie, S.; Wang, L. Source profiles, emission factors and associated contributions to secondary pollution of volatile organic compounds (VOCs) emitted from a local petroleum refinery in Shandong. *Environ. Pollut.* **2021**, *274*, 116589. https://doi.org/10.1016/j.envpol.2021.116589.
50. Rovira, J.; Nadal, M.; Schuhmacher, M.; Domingo, J.L. Environmental impact and human health risks of air pollutants near a large chemical/petrochemical complex: Case study in Tarragona, Spain. *Sci. Total Environ.* **2021**, *787*, 147550. https://doi.org/10.1016/j.scitotenv.2021.147550.
51. Bodor, K.; Szép, R.; Bodor, Z. Time series analysis of the air pollution around Ploiesti oil refining complex, one of the most polluted regions in Romania. *Sci. Rep.* **2022**, *12*, 11817. https://doi.org/10.1038/s41598-022-16015-7.

52. Sanda, M.; Dunea, D.; Iordache, S.; Pohoata, A.; Glod-Lendvai, A.-M.; Onutu, I.A. Three-Year Analysis of Toxic Benzene Levels and Associated Impact in Ploieşti City, Romania. *Toxics* **2023**, *11*, 748. https://doi.org/10.3390/toxics11090748.
53. Barbeş, S.-B.; Bărbulescu, A.; Barbeş, L. Assessing Benzene and TVOC Pollution and the Carcinogenic and Noncarcinogenic Risks to Workers in an Industrial Plant in Southeastern Romania. *Toxics* **2024**, *12*, 187.
54. Adebiyi, F.M. Air quality and management in petroleum refining industry: A review. *Environ. Chem. Ecotoxicol.* **2022**, *4*, 89–96.
55. Damian, C. Environmental pollution in the petroleum refining industry. *Ovidius Univ. Ann. Chem*. **2013**, *24*, 109–114.
56. Karbassi, A.R.; Abbasspour, M.; Sekhavatjou, M.S.; Ziviyar, F.; Saeedi, M. Potential for reducing air pollution from oil refineries. *Environ. Monit. Assess.* **2008**, *145*, 159–166.
57. Oil Refinery Emissions, A Cross-Border Impact Issue. Available online: https://kunakair.com/oil-refinery-emissions/ (accessed on 12 December 2024).
58. Refinery Air Emissions Management. Guidance Document for the Oil and Gas Industry. Available online: https://www.ipieca.org/resources/refinery-air-emissions-management (accessed on 12 December 2024).
59. Ritter, K.; Nordrum, S.; Shires, T.; Lev-On, M. Ensuring consistent greenhouse gas emissions estimates. *Chem. Eng. Prog*. **2005**, *101*, 30–37.
60. Bărbulescu, A.; Barbeş, L.; Nazzal, Y. New model for inorganic pollutants dissipation on the Northern part of the Romanian Black Sea coast. *Rom. J. Phys*. **2018**, *63*, 806.
61. Bărbulescu, A.; Barbeş, L. Characterization of the Concentrations of Volatile Organic Compounds in the Romanian Littoral using General Regression Neural Networks: A Case Study. *Anal. Lett*. **2016**, *49*, 387–399.