2025 IMEKO Joint Conference
TC8 - TC11 - TC25
Torino, Italy, September 14-17, 2025

Multivariate Analysis of Samples from the 2024 Manila Bay Oil Spill

Kim Christopher C. Aganda[1], Zamantha Nadir Z. Martin[1], Annabelle V. Briones[1], Rachel Zioness O. Diputado[2], Encar Alej D. Mendoza[2], Jaryl T. Jihad[2], and Hernando P. Bacosa[3]

[1] Department of Science and Technology - Industrial Technology Development Institute (DOST-ITDI), DOST Compound, Gen. Santos Ave., Bicutan, Taguig City, Philippines,
kccaganda@itdi.dost.gov.ph
[2] Philippine Coast Guard, 139 25th St., National Headquarters – Philippine Coast Guard, Port Area, Manila, Philippines
msif@coastguard.gov.ph
[3] Department of Science and Technology – National Research Council of the Philippines (DOST-NRCP), DOST Compound, Gen. Santos Ave., Bicutan, Taguig City, Philippines,
hernando.bacosa@msumain.edu.ph

***Abstract* – The 2024 Manila Bay oil spill was an impact of the monsoon rains enhanced by Super Typhoon Carina. Investigations utilizing chemical forensics were conducted to evaluate the impact of the accident. In this study, the application of multivariate analysis to aid the assessment was also explored.**

Samples collected at different points around the spill site and nearby coastal areas were prepared by extraction with dichloromethane followed by drying over sodium sulfate, while analysis was accomplished using gas chromatography – mass spectrometry (GC-MS) in Scan and SIM modes.

Common diagnostic biomarkers were used for comparison of the different oil samples. Principal component analysis and K-means clustering were also performed.

The study aims to provide an alternate approach to gaining insights into the properties of oil samples. The study also emphasizes the vital role of chemical measurements in making informed decisions and response efforts during environmental crises.

I. Introduction

The 2024 Manila Bay oil spill was a series of incidents involving three different tankers [1-3]. Due to the rains brought upon by Super Typhoon Carina at the time, it was feared that the oil slick will swiftly bring disastrous effects impacting different sectors, especially aquaculture [4-6]. It was especially crucial to determine the spread and impact of the spill given the prevalent weather conditions at the time.

Investigations and disaster relief efforts were swiftly initiated. A crucial part of these proceedings involved the identification of the type of oil/s as well as the source tanker/s of the spill. However, comparing multiple data sets with correlated variables simultaneously is a complex and tedious process.

Multivariate analysis (MVA) is a means of statistical treatment performed when several interrelated properties contribute to the variance of the sample set. This is particularly useful for complex samples, such as oil samples. Consequently, MVA was found effective in discriminating differences between similar types of various oil samples [7-13]. For this study, MVA was performed to determine supplementary information to support the investigation.

II. Methodology

A. Sample Preparation

Samples were obtained at multiple locations around the location of the accident and affected coastal areas. Around 250 mg of oil sample was extracted with 10 mL dichloromethane, then dried using sodium sulfate. Ultrasonication was also performed prior to drying samples with insoluble debris [14].

B. GC-MS Parameters

Gas chromatography was performed using the PerkinElmer Clarus 690 with the following temperature programming: (1) an initial temperature of 42°C for 1.5 minutes; (2) increase the temperature to 330°C at a ramp rate of 5.5 C°/min; and (3) hold at 330°C for 10 minutes. The equipment was fitted with an Elite 5-MS column with a length of 30 m, internal diameter of 250 um, and 0.25 um film thickness. Helium was used as carrier gas at 1.1 mL/min constant flow [15].

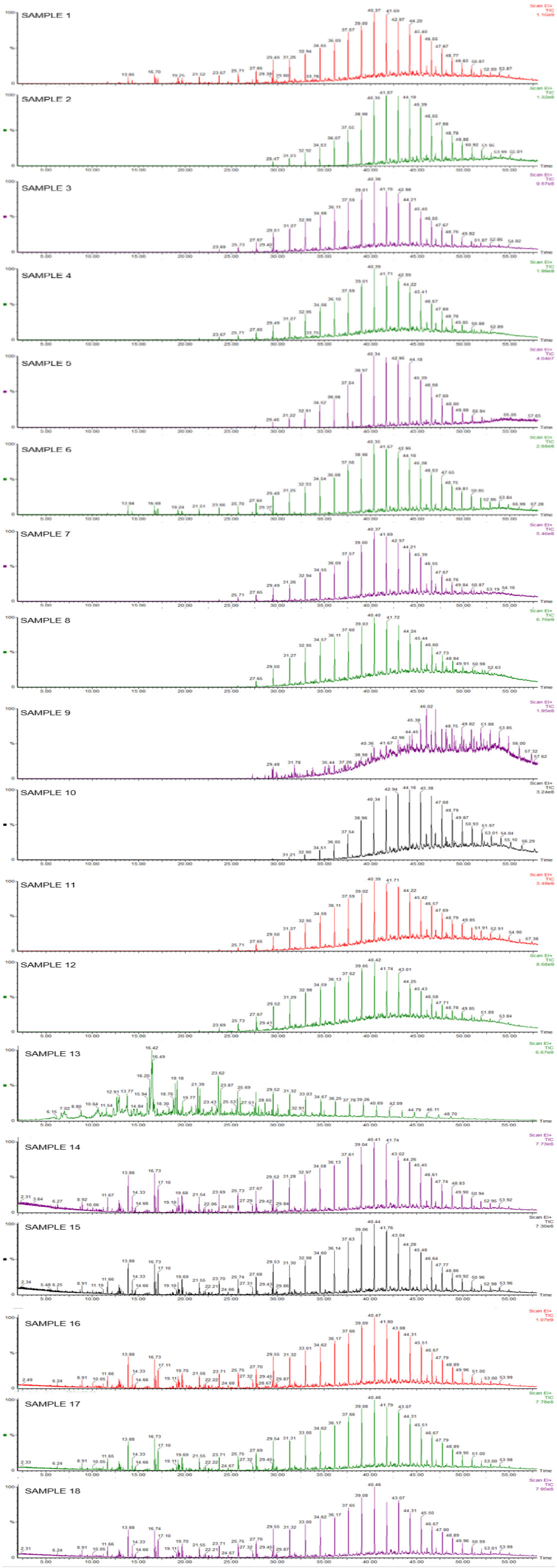

Fig. 1. Total ion chromatograms of the samples.

Mass spectrometry was accomplished with the PerkinElmer Clarus SQ8T equipped with a 70-eV electron impact ionization source. Data acquisition was done in Scan and selected ion monitoring (SIM) modes.

C. Comparison of diagnostic biomarkers

Common diagnostic biomarkers used in organic geochemistry and oil spill identification were used to compare the samples due to their specificity, diversity, and resistance to biodegradation and weathering. These include hopanes and tricyclic terpanes (m/z 191); steranes and diasteranes (m/z 217, 218); triaromatic steranes (m/z 231); and sesquiterpanes (m/z 123) [14][15]. The n-alkane profile (m/z 85) provides a general overview as well as weathering effects on the samples [14][15]. Diagnostic ratios were calculated for numerical comparison [15].

There are four different match levels for sample comparison [15]:

- Positive match (Ps)
  Comparison of chromatographic patterns and diagnostic ratios reveals minimal differences, which are either within the method variability or attributable to factors like weathering. This similarity indicates a high degree of scientific certainty that the samples match.
- Probable match (Pr)
  While chromatographic patterns and diagnostic ratios show differences that may not immediately indicate a definitive match, these discrepancies can be plausibly attributed to external factors such as weathering, mixing, or sample heterogeneity, suggesting a possible link between the samples.
- Inconclusive (I)
  A match is considered inconclusive if the differences in chromatographic patterns and diagnostic ratios cannot be ruled as either probable or non-matching, such as samples with low concentration.
- Non-match (N)
  Significant differences in chromatographic patterns and diagnostic ratios exceed method variability and cannot be attributed to external factors like weathering, contamination, or heterogeneity, indicating that the samples do not match with a high degree of scientific certainty.

D. Multivariate Analysis

Multivariate analysis was accomplished using principal component analysis (PCA) and K-means clustering. PCA is a means of dimension reduction in large data sets. Linear transformations of the data set are performed to obtain a smaller set of variables, called principal components, which represent most of the variance occurring in the data set [8,10,12,16]. K-means clustering is a means of iterative unsupervised clustering based on the distance between the centroids of the identified clusters. Data points are grouped into *k* clusters such that each data point belongs to the cluster with the closest cluster centroid [17-18].

The total ion chromatogram of each sample was extracted without further data cleanup. These were compiled in a single comma-separated values (CSV) file, then subjected to PCA, followed by K-means clustering.

The appropriate number of *k* clusters to describe the data set was calculated via the average silhouette method. The average silhouette method is a means of quality checking for the effectiveness of the clustering. The silhouette coefficient is a function of cluster cohesion (i.e. the average distance of a point to the other data points within the cluster) and cluster separation (i.e. the average distance of a given point to the other data points in the nearest cluster). The silhouette coefficient ranges from -1 to +1, where a value closer to +1 implies effective clustering at the given *k* value [18-21].

The overall results are then assessed in comparison to those obtained from the use of diagnostic ratios.

III. Results and Discussion

A. GC-MS Analysis

Figures 1 and 2 summarize, respectively, the total ion chromatograms and n-alkane profiles obtained. All samples were determined to be characteristic of heavy oil with different degrees of weathering.

Chromatograms for each of the diagnostic biomarkers were also acquired. These were assessed, via visual inspection [14][15] and numerical comparison using diagnostic ratios [15], in relation to Sample 1, which was obtained closest to the spill site and suspected to be the source oil. The interpretation of the comparison of the chromatographic patterns of Sample 1 with each sample is summarized in Table 1.

The chromatographic patterns of the biomarkers for Samples 3 and 4 were found to be positive matches with Sample 1. Overall, these were concluded to be positive matches with Sample 1. All three samples can be ascertained to have originated from the same source.

For Samples 7, 9, 14-18, biomarker comparison yielded probable matches with Sample 1. The differences may be attributed to external factors such as sample weathering and heterogeneity, among others. Thus, Samples 7, 9, and 14-18 were determined to be probable matches with Sample 1.

On the other hand, Samples 2, 6, 8, and 10-13 showed striking differences with Sample 1 which cannot be accounted for by external factors. Hence, these samples were ruled out as non-matches.

Lastly, Sample 4 was found to be inconclusive due to the lack of content viable for analysis.

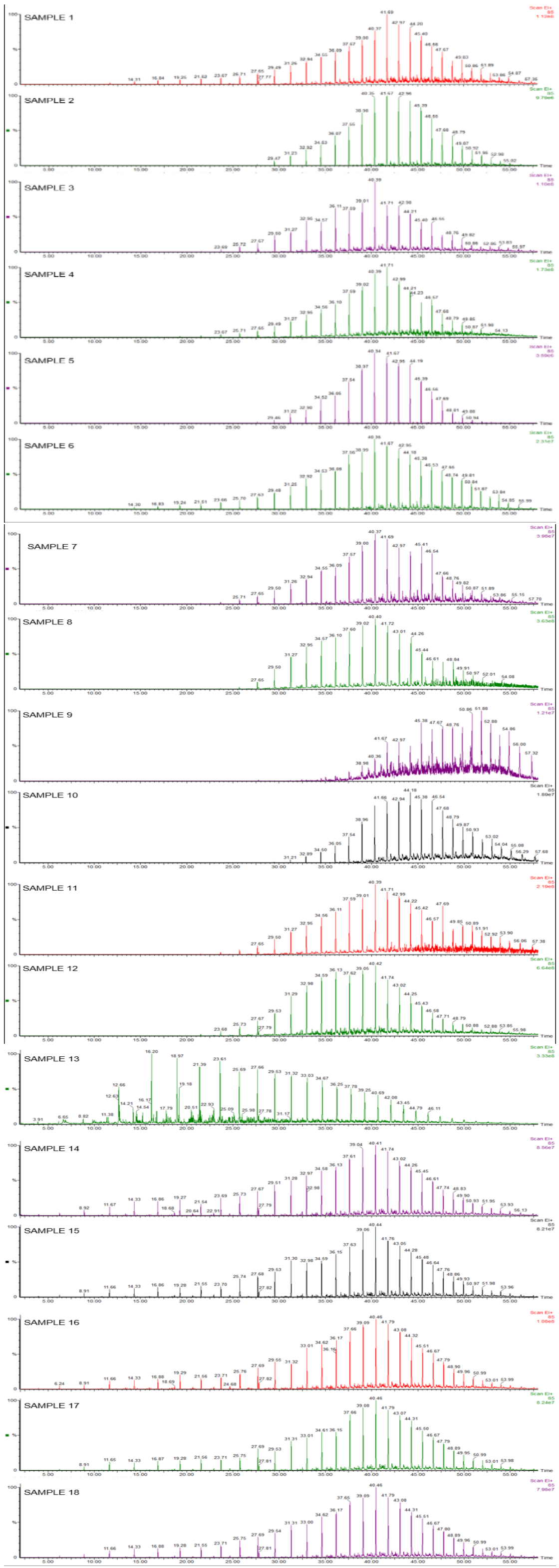

Fig. 2. Chromatograms of n-alkanes (m/z 85).

B. *Multivariate Analysis*

Principal component analysis was performed to transform the sample set to principal components by linear transformations, thus reducing the number of variables in the sample set while maximizing the variance among the components. The first three PCs (PC1, PC2, and PC3) account for 88.53% of the variance in the sample set as shown in Figure 3.

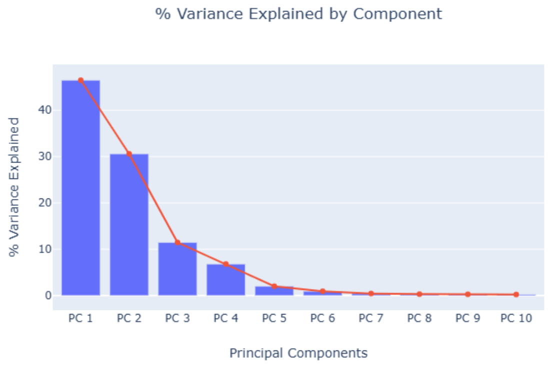

Fig. 3. Scree plot showing the variance as contributed by the calculated principal components.

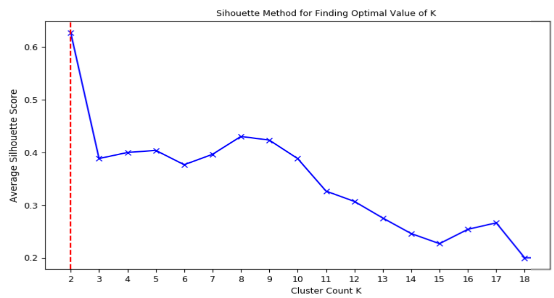

Fig. 4. Average silhouette coefficient results evaluated for different values of k.

With the K-means clustering, it was determined using the average silhouette method (as shown in Figure 4) that k=2 is sufficient to discriminate the sample sets.

Clusters were then determined using the value k=2 (shown in Figure 5). Upon closer inspection of the clusters, most of the samples in the set belong to one cluster (Cluster 1), with the rest found in another group (Cluster 2).

Cluster 1 contains most of the sample set. Almost all samples which were determined to be positive or probable match with the suspected source (Sample 1) are found in this cluster. This strongly suggests that the samples identified to be either positive or probable matches are very similar to Sample 1.

Cluster 2 contains only four samples: Samples 7, 11, 12, and 14. This implies that these samples share fewer similarities with Sample 1. Hence, these samples may not have originated from the same source. It should be noted that Samples 7 and 14 were classified as probable matches, while Samples 11 and 12 were determined as non-matches.

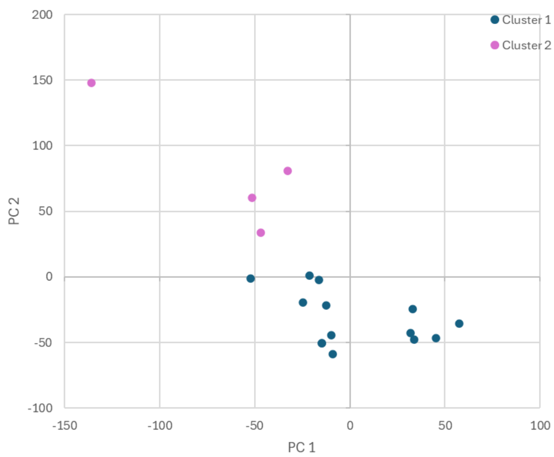

Fig. 5. PCA plot showing the identified clusters.

In general, the results from the multivariate analysis agree with the evaluation from the comparisons performed with the GC-MS results, corroborating with the findings as calculated using diagnostic ratios. Thus, multivariate analysis can be performed to aid investigations on oil spill fingerprinting.

Table 1. Interpretation of the comparison of the chromatographic patterns of Sample 1 to oil samples based on the five sets of biomarkers.

Sample	m/z 191	m/z 217	m/z 218	m/z 231	m/z 123	Over-all
Sample 2	I	N	N	N	N	N
Sample 3	Ps	Ps	Ps	Pr	Ps	Ps
Sample 4	Ps	Ps	Ps	Ps	Ps	Ps
Sample 5	N	I	I	I	I	I
Sample 6	I	N	N	N	N	N
Sample 7	Ps	Pr	Pr	Pr	Ps	Pr
Sample 8	N	N	N	N	N	N
Sample 9	Pr	Pr	Pr	Pr	Pr	Pr
Sample 10	N	N	N	N	Pr	N
Sample 11	N	N	N	N	N	N
Sample 12	N	N	N	N	N	N
Sample 13	N	N	N	I	N	N
Sample 14	Pr	Pr	I	I	Ps	Pr
Sample 15	Pr	Pr	I	I	Ps	Pr
Sample 16	Pr	Pr	Pr	Pr	Ps	Pr
Sample 17	Pr	Pr	I	I	Ps	Pr
Sample 18	Pr	Pr	I	I	Ps	Pr

IV. Conclusion

The results suggest that 2 out of the 17 samples show a positive match with the suspected oil source, while 7 were classified as probable matches, and 7 were found to be non-matches. One of the samples was ruled out as inconclusive due to low concentration.

Multivariate analysis showed that all the positive matches and most of the probable matches are grouped with the suspected source sample, suggesting similar origins. Furthermore, the similarities of the multivariate analysis results with the comparison assessment show that multivariate analysis can be a useful tool in oil spill impact surveys.

V. Acknowledgment

The research team gratefully acknowledges the support of the Philippine Coast Guard (PCG) for the provision of the oil samples, the Advanced Device and Materials Testing Laboratory (ADMATEL) of the Industrial Technology Development Institute (DOST-ITDI) for the facilities, and the Philippine Council for Industry, Energy, and Emerging Technology Research and Development (DOST-PCIEERD) for the resources extended for this endeavor.

References

[1] Bataan Oil Spill Bulletin No. 03, Series of 2024 07 August. Retrieved from https://www.bfar.da.gov.ph/2024/08/07/bataan-oil-spill-bulletin-no-03-series-of-2024-7-august-2024/

[2] Philippine Daily Inquirer. "Another fuel tanker sinks in Manila Bay." July 29, 2024. Retrieved from https://newsinfo.inquirer.net/1966839/another-fuel-tanker-sinks-in-manila-bay

[3] GMA News Online. "Third vessel found leaking hazardous material off Bataan." July 31, 2024. Retrieved from https://www.gmanetwork.com/news/topstories/regions/915357/third-vessel-found-leaking-hazardous-material-off-bataan/story/

[4] Bataan Oil Spill Bulletin No. 02, Series of 2024. Retrieved from https://www.bfar.da.gov.ph/2024/08/02/bataan-oil-spill-bulletin-no-02-series-of-2024/

[5] Reuters. "Philippine oil spill reaches fishing town, threatens livelihoods." July 29, 2024. Retrieved from https://www.reuters.com/world/asia-pacific/philippine-oil-spill-reaches-fishing-town-threatens-livelihoods-2024-07-29/

[6] Philippine Daily Inquirer. "Explainer: Manila Bay oil spills' harm on fishing, environment, public health." August 14, 2024. Retrieved from https://newsinfo.inquirer.net/1972664/explainer-manila-bay-oil-spills-harm-on-fishing-environment-public-health

[7] S. Wang, H. Zuo, C-N. Gao, J.-H. Wang, C. Li, and S. Wang, "Characterization of Differential Markers among Crude Oil Samples Using UPLC-QE-MS/MS and Multivariate Statistical Analysis," Energ. Fuel 37, 15 (2023).

[8] M. Mohammadi and M.K. Khorrami, "Application of robust principal component analysis-multivariate adaptive regression splines for the determination of API gravity in crude oil samples using ATR-FTIR spectroscopy," Arab. J. Chem. 16, 9 (2023).

[9] C.L. Ray, J.A. Gawenis, and C.M. Greenlief, "A New Method for Olive Oil Screening Using Multivariate Analysis of Proton NMR Spectra," Molecules 27 (1), 213 (2022).

[10] C.M.S. Sad, M. da Silva, F.D. dos Santos, L.B. Pereira, R.R.B. Corona, S.R.C. Silva, NA. Portela, E.V.R. Castro, P.R. Filgueiras, and V. Jr. Lacerda, "Multivariate data analysis applied in the evaluation of crude oil blends," Fuel 239, pp. 421-428 (2019).

[11] L.V. de Freitas, A.P.B.R. de Freitas, F.A.S. Marins, E.V. Veraszto, J.T.F. de Camargo, J.P. Davim, and M.B. Silva, "Contributions of Multivariate Statistics in Oil and Gas Industry." In: Multivariate Analysis in Management, Engineering and the Sciences. IntechOpen (2013).

[12] A. Silset, G.R. Flaten, H. Helness, E. Melin, G. Oye, and J. Soblom, "A Multivariate Analysis on the Influence of Indigenous Crude Oil Components on the Quality of Produced Water. Comparison Between Bench and Rig Scale Experiments," J. Disper. Sci. Technol. 31, 3 (2010).

[13] A. Hannisdal, PV. Hemmingsen, and J. Sjoblom, "Group-Type Analysis of Heavy Crude Oils Using Vibrational Spectroscopy in Combination with Multivariate Analysis," Ind. Eng. Chem. Res. 44, 5 (2005).

[14] ASTM D5739-06 (2013). Standard Practice for Oil Spill Source Identification by Gas Chromatography and Positive Ion Electron Impact Low Resolution Mass Spectrometry.

[15] PD CEN/TR 15522-2:2012. Oil spill identification – Waterborne petroleum and petroleum products – Part 2: Analytical methodology and interpretation of results based on GC-FID and GC-MS low resolution analyses.

[16] M. Greenacre, P.J.F. Groenen, T. Hastie, A.I. D'Enza, A. Markos, and E. Tuzhilina, "Principal component analysis," Nat. Rev. Methods Primers 2, 100 (2022).

[17] J. Wu, "Cluster Analysis and K-means Clustering: An Introduction." In: Advances in K-means Clustering. Springer Theses. Springer, Berlin, Heidelberg (2012).

[18] H. Belyadi and A. Haghighat, "Unsupervised machine learning: clustering algorithms." In: Machine Learning Guide for Oil and Gas Using Python. Elsevier (2021).

[19] F. Batool and C. Hennig, "Clustering with the Average Silhouette Width," Comput. Stat. Data Anal. 158, 107190 (2021).

[20] A. Subasi, "Clustering examples." In: Practical Machine Learning for Data Analysis Using Python. Elsevier (2020).

[21] J. Han, M. Kamber, and J. Pei, "Cluster Analysis: Basic Concepts and Methods." In: Data Mining: Concepts and Techniques (Third Edition). Elsevier (2012).