ISSN Online: 2821-1936

Transactions on Data Analysis in Social Science

Journal Homepage: https://transoscience.ir

Application of Positive Matrix Factorization (PMF) and Multivariate Statistical Techniques in Identifying and Managing Sources of Heavy Metal Pollutants in Sediments

F. Hedayatzadeh[1], N. Hassanzadeh[2], N. Bahramifar[3], A. Ildoromi[4,*]

[1] Ph.D. Student in Environmental Science, Environmental Group, Faculty of Natural Resources and Environment, Malayer University, Malayer, Iran
[2] Associate Professor, Watershed Management Group, Faculty of Natural Resources and Environment, Malayer University, Malayer, Iran
[3] Assistant Professor, Environmental Group, Faculty of Natural Resources and Environment, Malayer University, Malayer, Iran
[4] Associate Professor, Watershed Management Group, Faculty of Natural Resources and Environment, Malayer University, Malayer, Iran

ARTICLE INFO

Article History:
Received 17 May 2021
Received in revised form 20 July 2021
Accepted 30 August 2021
Available online 4 September 2021

Keywords:
Heavy metal pollutants, Source apportionment, Ecosystem management, Positive Matrix Factorization (PMF), Multivariate statistical techniques

ABSTRACT

Understanding and identifying various sources of water pollution and the processes affecting them are essential for achieving a comprehensive description of the quality of essential water resources. To this end, implementing a suitable network for monitoring water quality is crucial. Therefore, this study aims to examine the effectiveness of the Positive Matrix Factorization (PMF) model compared to multivariate statistical techniques. Additionally, the combined application of the PMF model with these methods in determining the contribution and management of heavy metal pollutants in aquatic sediments is investigated. Multivariate statistical analysis methods have proven effective in preparing and interpreting data on the quality of aquatic environments and determining the information available in them. However, they have some limitations. Thus, in this research, the application of Positive Matrix Factorization for sediment quality data, especially concerning heavy metals, is compared with multivariate statistical methods. The study also evaluates the status and extent of using this model in various environmental studies worldwide in recent years. Positive Matrix Factorization allows considering uncertain data and provides a positive constraint, leading to an environmentally interpretable result. The results of examining the applications of the PMF model in determining the contribution of various pollutants, including heavy metals, in different environmental sectors over the past two decades, indicate a significant increase in its usage in recent years compared to the past. Recent study results suggest that while Positive Matrix Factorization leads to a stronger understanding of the sources of pollution in the studied system compared to multivariate statistical methods, the combined application of the PMF model with other multivariate statistical methods for determining pollutant sources results in a more accurate and comprehensive analysis of pollutants, including heavy metals.

* Corresponding Author: ildoromi@gmail.com
Associate Professor, Watershed Management Group, Faculty of Natural Resources and Environment, Malayer University, Malayer, Iran

http://dx.doi.org/10.47176/TDASS.2021.136

1. INTRODUCTION

Aquatic ecosystems are the ultimate receptors of all pollutants, including heavy metals. Heavy metal pollution in aquatic environments is a growing global problem and has reached alarming levels [1]. Due to their high toxicity, stability, bioaccumulation potential, and transfer to the food chain, heavy metal pollution has become a fundamental issue in these systems [2]. Many studies have focused on sediments in aquatic and coastal areas as they act as important reservoirs for various toxic pollutants, including heavy metals discharged into aquatic environments [3-7]. Indeed, sediments can act as a reservoir of toxic compounds that continuously threaten the health and survival of aquatic organisms, potentially serving as a starting point for pollutant entry into the aquatic food web [8-9]. The quality of water resources, especially concerning heavy metals, can be influenced by various factors and processes, including human activities and natural phenomena. Understanding and identifying various sources of water pollution and the processes affecting them are essential for achieving a comprehensive description of the quality of essential water resources. To this end, implementing a suitable network for monitoring water quality and identifying potential sources of heavy metals in sediments is a fundamental requirement to prevent further pollution of aquatic environments [10-11].

However, the generated databases are large and complex, requiring powerful analytical tools for their analysis. Multivariate statistical techniques are widely used to assess temporal and spatial variations and interpret complex data from water resources [12-18]. Multivariate statistical analysis methods are an approach to reduce the dimensions of data, providing a simplified representation of the most important factors. These methods are generally used to determine the structure of data and provide qualitative information about potential pollution sources. However, they alone cannot quantitatively determine the contribution of identified pollution sources for each variable. Receiver-based models can be used for this purpose.

Receiver-based models, as a type of source identification model, have been used to identify various pollution sources and calculate their contribution or role in sediment pollution [19]. These models can effectively reduce the dimensions of grouped data and variables based on their common characteristics [10, 20]. One of the receiver model methods is Positive Matrix Factorization (PMF), which has been used to determine the quantitative contribution of different pollutant sources in surface sediment in aquatic environments [21-24]. As an advanced source apportionment model, PMF is used to determine the quantitative contribution of pollutant sources and has provided useful information in many studies [25-26]. Since obtaining quantitative information about the contribution of various sources for each heavy metal is a prerequisite for formulating effective strategies to control the release of trace elements [19, 25, 27], this study aims to investigate the effectiveness of one of the widely used receiver models, Positive Matrix Factorization (PMF), and the combination of this model with multivariate statistical methods and spatial statistics in determining the contribution of heavy metal pollutant sources in sediments of aquatic environments, compared to the application of some common multivariate statistical methods alone.

2. LITERATURE REVIEW

Various methods such as Spatial Deviation (SD), Correlation Analysis (CA), Enrichment Factor (EF), Principal Component Analysis (PCA), Hierarchical Cluster Analysis (HCA), Factor Analysis (FA), and Geographic Information System (GIS) have been used to find heavy metal sources in the environment. These approaches can quickly identify the common features of new components through classification or dimensionality reduction. However, accurate tracking and calculation of the contribution of various sources using these methods are not obtained. Positive Matrix Factorization (PMF) recommended by the EPA is an ideal receptor model for identifying and determining sources. It decomposes the primary dataset into a factor matrix and a profile group to calculate the contribution and distribution of sources. The practical application of combining the PMF model with other multivariate statistical methods and GIS can be an effective tool for identifying pollutant sources. However, few quantitative studies compare and combine different source identification approaches in polluted areas. In this section, some studies conducted in this field are mentioned:

Huang et al. (2018) conducted a study to assess pollution and identify the sources of heavy metals in the sediments of the Nantong River in eastern China. In this research, four potential main sources of heavy metals were identified using Principal Component Analysis (PCA) and Cluster Analysis (CA) methods [28].

Zhang et al. (2018) conducted a study to identify the sources of arsenic and heavy metals in agricultural soils around the industrial city of Hunan, located in the northern Yangtze River in China. They utilized the PMF model, PCA, and GIS. The results of this study also indicated that a combination of Positive Matrix Factorization (PMF) with multivariate statistical models and spatial statistics is highly effective in determining and allocating the contribution of heavy metal sources [29].

Dong et al. (2019) evaluated and identified the sources of heavy metals in agricultural soils in the Baiyin region of northwestern China. The identification and determination of heavy metal sources in this study were performed using various methods, including Spatial Deviation (SD), Cluster Analysis (CA), Enrichment Factor (EF), Principal Component Analysis (PCA), Geographic Information System (GIS), and Positive Matrix Factorization (PMF). The results showed that the combined applications of GIS, PCA, and PMF for determining heavy metal sources are accurate, practical, and effective [30].

Boroumandi et al. (2019) also assessed soil pollution sources using a combination of multivariate statistical methods (PMF, PCA, and CA) and spatial statistics in the Zanjan watershed in Iran [31]. Their study results indicated that the PMF method provided more acceptable results compared to other common multivariate statistical methods. Zhang et al. (2020) investigated and determined the sources of heavy metal pollution in the rhizosphere soil of irrigated farms by the Yellow River in China, using the PMF model along with multiple statistical methods (Coefficient of Variation (CV), Correlation Analysis (CA), Enrichment Factor (EF), and Principal Component Analysis (PCA)) [32]. The results of this study also demonstrated that the combination of common methods with the PMF model leads to more accurate and comprehensive results in identifying pollutant sources.

3. MATERIALS AND METHODS

In this section, some common multivariate statistical techniques used for determining the sources of pollutant emissions, including heavy metals, and the Positive Matrix Factorization (PMF) model, are discussed.

3.1. Multivariate Statistical Techniques

Multivariate statistical methods, including Principal Component Analysis (PCA), Factor Analysis (FA), and Hierarchical Cluster Analysis (HCA), are significantly employed for identifying the sources and evaluating the status of heavy metal pollution in aquatic sediments [33].

3.1.1. Principal Component Analysis (PCA)

Principal Component Analysis is a mathematical tool that utilizes orthogonal transformations to convert a set of potentially correlated variables, called principal components (PCs). The primary idea of PCA is to reduce the dimensionality of a dataset consisting of numerous variables while retaining the essential information to a minimum loss [34]. The principal component (PC) can be expressed as follows:

$$Z_{ij} = \alpha_{i1}X_{1j} + \alpha_{i2}X_{2j} + \alpha_{i3}X_{3j} + \cdots + \alpha_{im}X_{mj} \tag{1}$$

Where Z is the score of the factors, α is the factor loading, X is the measured value of the variable, i is the number of factors, j is the number of samples, and m is the total number of variables. In PCA, PCs with eigenvalues greater than 1 are generally considered and contain the most diversity in the original dataset. Since principal loads may not be easily interpretable and are usually rotated until a "simple structure" is achieved, it means that each variable has very high factor loads (close to 1) in one of the PCs and very low factor loads (close to zero) in the other PCs.

3.1.2. Factor Analysis (FA)

Factor Analysis is a multivariate statistical analysis aiming to find relationships between parameters by reducing the number of measured parameters to a limited number of factors that can explain most of the data variance without additional information and noise in the data [35]. This process, known as FA, follows PCA, and the newly constructed variables are also referred to as Vareimax (VFs). FA can be expressed as follows:

$$Z_{ij} = \alpha_{f1} f_{1i} + \alpha_{f2} f_{2i} + \alpha_{f3} f_{3i} + \cdots + \alpha_{fm} f_{mi} + e_{fi} \quad (2)$$

Where Z is the factor score, α is the factor loading, *f* is the factor score, e is the error or other source of variation, i is the number of samples, and m is the number of variables. The KMO test and Bartlett's test are used when the sample k from the population has equal variance. The KMO statistic ranges between 0 and 1, and Kaiser (1974) recommends values above 0.5 as acceptable. Additionally, values between 0.5 and 0.7 are mediocre, values between 0.7 and 0.8 are good, values between 0.8 and 0.9 are excellent, and values above 0.9 are outstanding [37-36]. The SPSS software is used for conducting FA, and correlation matrices are used for factor analysis [38].

3.1.3. Hierarchical Cluster Analysis (HCA)

HCA is another multivariate statistical technique that classifies elements based on similar features [40-39]. Essentially, HCA is based on hierarchical grouping that starts with the most similar traits and gradually develops higher-level groups. The process of forming and merging groups is repeated until a single group containing all samples is obtained. The result can be displayed as a dendrogram, providing a graphical summary of the clustering process. HCA uses the Ward method with Euclidean distance as the similarity measure [41].

3.2. Positive Matrix Factorization (PMF) Model

Positive Matrix Factorization (PMF) [42] is also a multivariate receptor model designed specifically for identifying and apportioning source contributions, particularly for addressing environmental data and managing their uncertainty and distribution. PMF is suitable for environmental data as it considers uncertainties in the analysis often associated with environmental sample measurements and ensures that all values in the profiles and source contributions are positive, leading to a more realistic representation of a system compared to other multivariate methods such as PCA [43]. Two input files are entered into the PMF model: one file containing the concentrations of pollutants under investigation and another file containing uncertainty values. The quantitative determination of heavy metal sources using the PMF model is calculated according to the following equations:

$$X=GF+E \quad (3)$$

where X ($m \times n$) is the data matrix containing measurements of *m* factors in *n* samples, G ($m \times p$) is the factor or source profile matrix, F ($p \times n$) is the factor score or source contribution matrix, and E ($m \times n$) is an error matrix.

$$X_{ij} = \int_{k=1}^{p} g_{ik} f_{kj} + e_{ij} \quad (4)$$

where X_{ij} is the concentration of species *j* in sample *i* $(mg/kg)^{-1}$, g_{ik} is the contribution of source *k* $(mg/kg)^{-1}$, f_{kj} is the amount of species *j* from source *k*, e_{ij} is the error value, and *p* is the number of source factors.
The uncertainty data file can be based on both observations and equations. The uncertainty file is calculated based on equations for specific parameters for each sample through equations provided by the EPA [43]. Uncertainty is calculated based on the method detection limit (MDL) constraint, and the percentage error is determined by standard reference materials. If the measured concentration is less than the MDL for that particular species, uncertainty data is calculated as follows:

$$\text{Unc}= \frac{5}{6} \times \text{MDL} \quad (5)$$

and if the concentration is greater than the provided MDL, the calculation is based on a fraction of the concentration and MDL:

$$\mathrm{Unc} = \sqrt{(\text{error fraction} \times \mathrm{C})^2 + (0.5 \ \times \mathrm{MDL})^2} \tag{6}$$

4. RESULTS AND DISCUSSION

4.1. Application of Multivariate Statistical Techniques and PMF Model in Identifying the Sources of Pollutants, Including Heavy Metals, in the Environment

Multivariate statistical techniques, including Factor Analysis (FA), Principal Component Analysis (PCA), and Hierarchical Cluster Analysis (HCA), are used for the analysis of complex multidimensional water quality datasets. These techniques aid in identifying pollution sources in surface waters and enhance understanding of the ecological status and water quality of the studied systems [45-50]. Multivariate statistical methods are employed to assess spatiotemporal patterns related to human and natural factors [51-52]. Factor Analysis (FA) and Principal Component Analysis (PCA) are widely used for determining the contribution and distribution of water quality sources in surface waters [52-54].

However, allocating the share of water quality sources based on water quality data matrices from monitoring stations using the PMF model is quite rare. This is while PMF is a novel technique that has brought some advantages over other multivariate techniques such as FA, PCA, and HCA, including the inclusion and calculation of uncertainty. For this reason, PMF has demonstrated superiority in many environmental applications compared to other multivariate statistical methods.

As environmental data matrices typically have data points below the detection limit and missing values [55], estimating uncertainty in PMF modeling allows for better correction of values below the detection limit and missing values [56]. Therefore, employing and estimating uncertainty values in the PMF model leads to effective utilization of this model in various environmental systems to determine the contribution of different pollutant sources. Figure 1 provides a general schematic of the applications of the PMF model in determining the contribution of sources such as aerosol particles, volatile organic compounds (VOCs), heavy metals (HMs), aromatic hydrocarbons (PAHs), and other pollutants in three different environmental compartments: air, water, and soil.

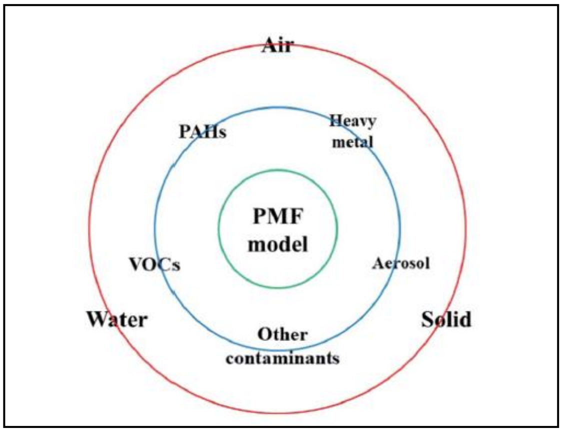

Fig. 1. General Schematic of the Application of the PMF Model in the Environment

The successful applications of PMF are primarily reported in studies related to the allocation of suspended particle sources in the environment, with numerous investigations available in this field [57-58]. Additionally, Henry and Christensen (2010) introduced the use of the PMF model in aquatic environments to reduce uncertainty weight existing in datasets, aiming for more reasonable results. In addition to air pollution studies, the application of this technique has expanded to sediments and aquatic environments to identify the contribution of various sources to the settling of moist pollutants such as TOC, NH4-N, NO3-N, Mg, K, and Ca as source tracers [59]. Moreover, the use of this model for the allocation of various other pollutant sources in aquatic environments, such as TP, NH3, BOD, PAHs, PCBs, and HMs, has recently extended [60-61].

Over the past two decades, PMF has been widely used in studies related to air pollution. Particularly, PMF has been extensively applied to examine the distribution of suspended particles in the air in several countries [62-67]. Only in recent years, some researchers have employed PMF in datasets related to various environmental areas, including sediment in aquatic environments, to achieve a more realistic understanding of the effective sources in these systems and the better efficiency of this model compared to other multivariate statistical methods. Recent studies on determining pollutant sources in sediment in aquatic environments using the PMF model include:

- Comero et al. (2011) identified four interpretable factors related to the mineralogical/chemical characteristics of lake sediments in the Alpine lakes dataset in Italy using the PMF model.
- Pekey and Dogan (2013) investigated the application of Positive Matrix Factorization (PMF) in assessing the distribution of heavy metal sources in the sediments of the western Gulf of İzmit in Turkey, comparing it with a previous factor analysis.
- González-Macías et al. (2014) used Positive Matrix Factorization (PMF) to examine and identify the contribution of heavy metal sources in coastal sediments of the Pacific Ocean in Mexico.
- Li et al. (2015) utilized the Positive Matrix Factorization model to determine the contribution of various sources to the surface water quality of the Daliao River watershed in northeastern China.
- Haji Gholizadeh et al. (2016) used APCS-MLR and PMF receptor models to determine pollution sources in three rivers in southern Florida.
- Zanotti et al. (2019) conducted a study comparing the effectiveness of the PMF model in identifying characteristics and allocating pollutant sources in surface and groundwater of the Oglio River watershed with the multivariate statistical method Hierarchical Cluster Analysis (HCA).
- Jiang et al. (2019) used a combined method of hierarchical cluster analysis (HCA) and Positive Matrix Factorization (PMF) to identify non-point sources (NPs) for the Huaihe River basin in China.
- Salim et al. (2019) employed PCA-MLR and PMF receptor models to identify and allocate pollution sources in watersheds and sub-watersheds in South Korea.

Despite the use of the PMF model in studies related to pollutants such as heavy metals in aquatic environments, its overall application in determining pollutant sources in aquatic systems is much less compared to atmospheric pollutants. In this regard, the lower adoption of the PMF model in water research, compared to its use in air pollution studies, is attributed to the less suitability of water quality data to meet the requirements of this method. Some limitations that can potentially restrict the successful application of PMF include (1) PMF requires data reported as concentrations, while some common water measurements have different units (e.g., pH, electrical conductivity, redox potential, isotope analysis), making them incompatible directly in a PMF model; (2) an adequate monitoring network, especially for capturing system variables and pollution sources, is crucial for obtaining a comprehensive representation of the targeted system with PMF analysis; and (3) in cases where a single source exists, multivariate statistical analysis like PMF may not be suitable [72]. Despite these limitations, the potential benefits of PMF (such as including analytical uncertainties and positive constraints) can be useful when examining water sources and water quality datasets to derive meaningful water quality characteristics.

According to a meta-analysis conducted by Sun et al. (2020) on the use of the PMF model in determining the contribution of various pollutant sources, including heavy metals, in different environmental compartments during the past two centuries (2002-2018) globally, their study results indicate an increasing use of this model in recent years compared to the past. Based on the graph presented in Figure (2), no studies on the use of the PMF model in determining the contribution of heavy metal sources in the environment were conducted or reported before 2008. However, a significant increase in the frequency of using this model for heavy metals in various studies, especially in 2018, is clearly observable. This finding, considering the advantages of using this model compared to other multivariate statistical techniques, is promising for gaining a stronger understanding of various pollutant sources, especially heavy metals in the studied systems, over the past few years.

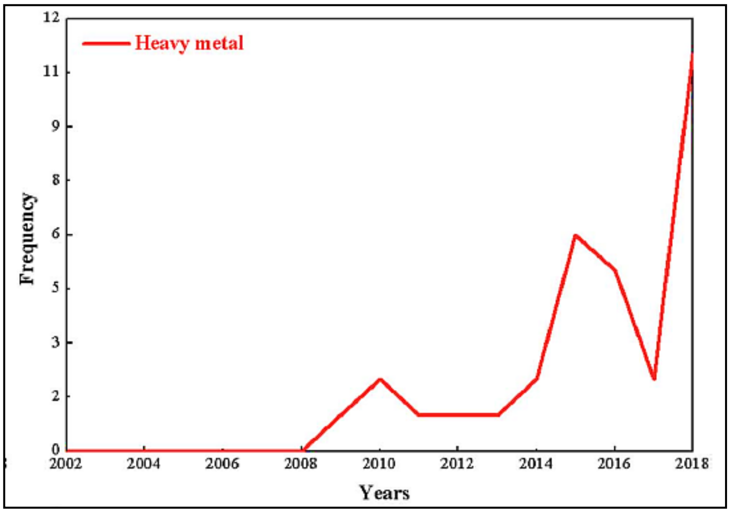

Fig. 2. Frequency of PMF Model Usage in Recent Years for Heavy Metals in Various Studies (Sun et al., 2020)

Additionally, in this study [75], a comparison of the percentage of PMF model usage for determining the contribution of pollutants such as PAHs, HMs, and VOCs across different continents worldwide from 2002 to 2018 has been presented. The results show that the highest percentage of PMF model usage for all mentioned pollutants was in Asia, followed by Africa and Europe, compared to other continents (Figure 3). In Asia, the highest percentage of PMF model usage was for determining the contribution of heavy metal sources, followed by aromatic hydrocarbons and to a lesser extent fugitive organic compounds. In Europe, there was almost uniform usage of this model in determining the contribution of sources for these pollutants. However, in Africa, in contrast to Asia, the highest percentage of PMF model usage was for determining the contribution of fugitive organic compounds, with the lowest percentage for determining the contribution of heavy metal sources. Therefore, based on these results, it can be inferred that the highest percentage of PMF model usage has been in Asia, where the application of this model has been more focused on determining the contribution of heavy metal sources compared to other pollutants.

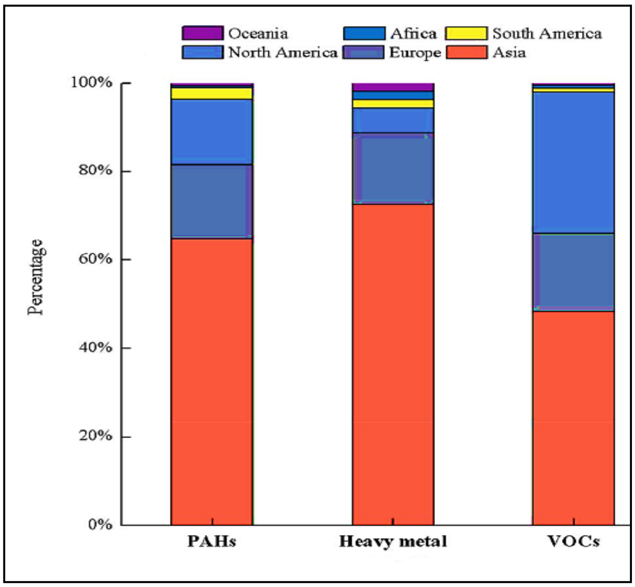

Fig. 3. Percentage of PMF Model Usage for Determining the Contribution of PAHs, HMs, and VOCs Globally from 2002 to 2018 (Sun et al., 2020)

4.2. Using a Combined Approach of PMF Model with Other Multivariate Statistical Methods and GIS

Recent studies have shown that the practical combination of the PMF model with other multivariate statistical methods and GIS can serve as an effective tool for identifying pollutant sources [29, 30, 32]. The reason behind this is that each of these methods has its limitations, and by combining them, more accurate and comprehensive results can be obtained in the field of identifying pollutant sources. For example, multivariate statistical methods such as PCA, FA, or CA are commonly used to approximate pollutants like heavy metals, and spatial analysis through GIS visually reflects the local situation, confirming whether PCA results are reasonable or not. On the other hand, the PMF model is used for a more in-depth analysis of the types of pollutant sources and their quantitative contributions based on previous results. Although the PMF method can qualitatively identify various pollutant sources and provide their quantitative contributions, when used alone, the qualitative analysis of pollutant sources relies on summarizing and screening previous data, which may lead to mental biases in interpreting PMF analysis results.

However, if PMF analysis is combined with other pollution assessment methods (Coefficient of Variation (CV), Correlation Analysis (CA), Enrichment Factor (EF), Factor Analysis (FA), Principal Component Analysis (PCA), and even Geographic Information System (GIS)), the limitations of PMF model analysis can be mitigated [30]. Therefore, it can be concluded that the application of each of these methods individually does not lead to a comprehensive and systematic analysis of pollutant sources, including heavy metals. However, if conventional methods are combined with the PMF model for determining pollutant sources, the results will be more accurate and comprehensive. In conclusion, the combination of these methods is highly suitable for the analysis of pollutant sources.

5. CONCLUSION

Based on this study, it can be acknowledged that despite the increased application of the PMF model in environmental studies in recent years compared to the past, and the superiority of PMF in many environmental applications over multivariate statistical methods, the practical applications of the PMF model for determining the sources of pollutants in surface water from water quality monitoring networks are limited. On the other hand, the conventional applications of FA and PCA for identifying water pollution sources have been widely reported. However, considering that the existence of data below the detection limit (BDL) and some missing data points in environmental datasets is common, and also due to the PMF model's ability to handle weighted data points effectively, controlling BDL values and missing data, its use leads to a stronger understanding of the studied system compared to the application of multivariate statistical methods alone. In other words, positive matrix factorization (PMF) analysis can be a useful method for determining environmental characteristics and quality, providing a deeper understanding of the dominating phenomena in the environment. However, as mentioned, when this method is used alone, the qualitative analysis of pollution sources relies solely on summarizing and screening previous data, shaping the mental image of PMF analysis results. Nevertheless, if PMF analysis is combined with other multivariate statistical methods, the limitations of the PMF model can be mitigated. Therefore, it can be concluded that the combined application of the PMF model with other multivariate statistical methods for identifying pollution sources, including heavy metals, results in a comprehensive and systematic analysis of pollution sources, leading to more accurate and comprehensive outcomes.

Transparency Statement

The data supporting this study are available upon reasonable request to the corresponding author, subject to ethical and confidentiality considerations.

Acknowledgments

We would like to express our gratitude to all individuals who contributed to this project.

Declaration of Interest

The authors declare that they have no competing interests.

Funding

This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

REFRENCESS

[1] Ogoyi, D. O., Mwita, C. J., Nguu, E. K., & Shiundu, P. M. (2011). Determination of heavy metal content in water, sediment and microalgae from Lake Victoria, East Africa. The Open Environmental Engineering Journal, 4, 156–161.

[2] Li, X., Wang, Y., Li, B., Feng, C., Chen, Y., & Shen, Z. (2013). Distribution and speciation of heavy metals in surface sediments from the Yangtze estuary and coastal areas. Environmental Earth Sciences, 69, 1537–1547.

[3] Chen, Y., Liu, R., Sun, C., Zhang, P., Feng, C., & Shen, Z. (2012). Spatial and temporal variations in nitrogen and phosphorous nutrients in the Yangtze River Estuary. Marine Pollution Bulletin, 64, 2083–2089.

[4] Liu, R. M., Chen, Y. X., Sun, C. C., Zhang, P. P., Wang, J. W., Yu, W. W., & Shen, Z. Y. (2014). Uncertainty analysis of total phosphorus spatial–temporal variations in the Yangtze River Estuary using different interpolation methods. Marine Pollution Bulletin, 86, 68–75.

[5] Battarbee, R. W., Turner, S., Yang, H., Rose, N. L., Smyntek, P. M., Reimer, P. J., Oldfield, F., Jones, V. J., Flower, R. J., Roe, K., Shilland, E., & Blaauw, M. (2015). Air pollutant contamination and acidification of surface waters in the North York Moors, UK: Multiproxy evidence from the sediments of a moorland pool. The Holocene, 25, 226–237.

[6] Syakti, A. D., Demelas, C., Hidayati, N. V., Rakasiwi, G., Vassalo, L., Kumar, N., Prudent, P., & Doumenq, P. (2015). Heavy metal concentrations in natural and human-impacted sediments of Segara Anakan Lagoon, Indonesia. Environmental Monitoring and Assessment, 187, 4079.

[7] Wang, J. W., Liu, R. M., Wang, H. T., Yu, W. W., Xu, F., & Shen, Z. Y. (2015). Identification and apportionment of hazardous elements in the sediments in the Yangtze River estuary. Environmental Science and Pollution Research, 22, 20215–20225.

[8] Ding, Z., & Hu, X. (2013). Assessment of As and metallic element contamination of riverine sediments from Yangze River in Nanjing reach, China. Fresenius Environmental Bulletin, 22, 824–831.

[9] Duan, X., Li, Y., Li, X., Li, M., & Zhang, D. (2013). Distributions and sources of polychlorinated biphenyls in the coastal East China Sea sediments. Science of the Total Environment, 463, 894–903.

[10] Gonzalez-Macias, C., Sanchez-Reyna, G., Salazar-Coria, L., & Schifter, I. (2014). Application of the positive matrix factorization approach to identify heavy metal sources in sediments: A case study on the Mexican Pacific Coast. Environmental Monitoring and Assessment, 186, 307–324.

[11] Yuan, X., Zhang, L., Li, J., Wang, C., & Ji, J. (2014). Sediment properties and heavy metal pollution assessment in the river, estuary and lake environments of a fluvial plain, China. Catena, 119, 52–60.

[12] Bhat, S. A., Meraj, G., Yaseen, S., & Pandit, A. K. (2014). Statistical assessment of water quality parameters for pollution source identification in Sukhnag stream: An inflow stream of Lake Wular (Ramsar site), Kashmir Himalaya. Journal of Ecosystems, 1–19.

[13] Duan, W., He, B., Nover, D., Yang, G., Chen, W., Meng, H., Zou, S., & Liu, C. (2016). Water quality assessment and pollution source identification of the eastern Poyang Lake Basin using multivariate statistical methods. Sustainability, 8(2), 133.

[14] Juahir, H., Zain, S. M., Yusoff, M. K., Hanidza, T. I. T., Armi, A. S. M., Toriman, M. E., & Mokhtar, M. (2011). Spatial water quality assessment of Langat River basin (Malaysia) using environmetric techniques. Environmental Monitoring and Assessment, 173, 625–641.

[15] Kumar, A. S., Reddy, A. M., Srinivas, L., & Reddy, P. M. (2014). Assessment of surface water quality in Hyderabad Lakes by using multivariate statistical techniques, Hyderabad-India. Environmental Pollution, 4(2), 14–23.

[16] Muangthong, S., & Shrestha, S. (2015). Assessment of surface water quality using multivariate statistical techniques: Case study of the Nampong River and Songkhram River, Thailand. Environmental Monitoring and Assessment, 187(9), 548.

[17] Mustapha, A., & Aris, A. Z. (2012). Spatial aspects of surface water quality in the Jakara Basin, Nigeria using chemometric analysis. Journal of Environmental Science and Health, Part A, 47, 1455–1465.

[18] Xiao, M., Bao, F., Wang, S., & Cui, F. (2016). Water quality assessment of the Huaihe River segment of Bengbu (China) using multivariate statistical techniques. Water Resources, 43(1), 166–176.

[19] Lin, Y., Chang-Chien, G., Chiang, P., Chen, W., & Lin, Y. (2013). Multivariate analysis of heavy metal contaminations in seawater and sediments from a heavily industrialized harbor in Southern Taiwan. Marine Pollution Bulletin, 76, 266–275.

[20] Qu, M. K., Li, W. D., Zhang, C. R., Wang, S. Q., Yang, Y., & He, L. Y. (2013). Source apportionment of heavy metals in soils using multivariate statistics and geostatistics. Pedosphere, 23(4), 437–444.

[21] Comero, S., Locoro, G., Free, G., Vaccaro, S., De Capitani, L., & Gawlik, B. M. (2011). Characterisation of Alpine lake sediments using multivariate statistical techniques. *Chemometrics and Intelligent Laboratory Systems, 107*, 24–30.

[22] Rahman, S. A., Hamzah, M. S., Wood, A. K., Elias, M. S., Salim, N. A. A., & Sanuri, E. (2011). Sources apportionment of fine and coarse aerosol in Klang Valley, Kuala Lumpur using positive matrix factorization. *Atmospheric Pollution Research, 2*, 197–206.

[23] Malandrino, M., Di Martino, M., Giacomino, A., Geobaldo, F., Berto, S., Grosa, M. M., & Abollino, O. (2013). Temporal trends of elements in Turin (Italy) atmospheric particulate matter from 1976 to 2001. *Chemosphere, 90*, 2578–2588.

[24] Tian, Y., Shi, G., Han, S., Zhang, Y., Feng, Y., Liu, G., Gao, L., Wu, J., & Zhu, T. (2013). Vertical characteristics of levels and potential sources of water-soluble ions in PM10 in a Chinese megacity. *Science of the Total Environment, 447*, 1–9.

[25] Xu, J., Peng, X., Guo, C. S., Xu, J., Lin, H. X., Shi, G. L., & Tysklind, M. (2016). Sediment PAH source apportionment in the Liaohe River using the ME2 approach: A comparison to the PMF model. *Science of the Total Environment, 553*, 164–171.

[26] Xue, J., Zhi, Y., Yang, L., Shi, J., Zeng, L., & Wu, L. (2014). Positive matrix factorization as source apportionment of soil lead and cadmium around a battery plant (Changxing County, China). *Environmental Science and Pollution Research, 21*, 7698–7707.

[27] Panda, U. C., Rath, P., Bramha, S., & Sahu, K. C. (2010). Application of factor analysis in geochemical speciation of heavy metals in the sediments of a lake system-Chilika (India): A case study. *Journal of Coastal Research, 26*, 860–868.

[28] Huang, S., Tu, J., Jin, Y., Hua, M., Wu, H., Xu, W., Yang, Y., Wang, H., Su, Y., & Cai, L. (2018).

Contamination assessment and source identification of heavy metals in river sediments in Nantong, Eastern China. *International Journal of Environmental Research, 12*, 373–389.

[29] Zhang, X., Wei, S., Sun, Q., Wadood, S. A., & Guo, B. (2018). Source identification and spatial distribution of arsenic and heavy metals in agricultural soil around Hunan industrial estate by positive matrix factorization model, principal components analysis, and geostatistical analysis. *Ecotoxicology and Environmental Safety, 159*, 354–362.

[30] Dong, B., Zhang, R., Gan, Y., Cai, L., Freidenreich, A., Wang, K., Guo, T., & Wang, H. (2019). Multiple methods for the identification of heavy metal sources in cropland soils from a resource-based region. *Science of the Total Environment, 651*, 3127–3138.

[31] Boroumandi, M., Khamehchiyan, M., Nikoudel, M. R., & Mohammadzadeh, M. (2019). Evaluation of soil pollution sources using multivariate analysis combined with geostatistical methods in Zanjan Basin, Iran. *Geopersia, 9*(2), 293–304.

[32] Zhang, M., Wang, X., Liu, C., Lu, J., Qin, Y., Mo, Y., Xiao, P., & Liu, Y. (2020). Identification of the heavy metal pollution sources in the rhizosphere soil of farmland irrigated by the Yellow River using PMF analysis combined with multiple analysis methods: Using Zhongwei city, Ningxia, as an example. *Environmental Science and Pollution Research, 27*(14), 16203–16214.

[33] Chai, L., Li, H., Yang, Z., Min, X., Liao, Q., Liu, Y., Men, S., Yan, Y., & Xu, J. (2017). Heavy metals and metalloids in the surface sediments of the Xiangjiang River, Hunan, China: Distribution, contamination, and ecological risk assessment. *Environmental Science and Pollution Research, 24*, 874–885.

[34] Helena, B., Pardo, R., Vega, M., Barrado, E., Fernandez, J. M., & Fernandez, L. (2000). Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis. *Water Research, 34*, 807–816.

[35] Papaioannou, A., Mavridou, A., Hadjichristodoulou, C., Papastergiou, P., Pappa, O., Dovriki, E., & Rigas, I. (2009). Application of multivariate statistical methods for groundwater physicochemical and biological quality assessment in the context of public health. *Environmental Monitoring and Assessment, 170*, 87–97.

[36] Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39*, 31–36.

[37] Hutcheson, G., & Sofroniou, N. (1999). *The multivariate social scientist: Introductory statistics using generalized linear models.* London, UK: Sage.

[38] Todeschini, R. (1998). *Introduzione alla chemiometria.* Napoli, Italy: Edises.

[39] Chen, P., Li, L., & Zhang, H. (2015). Spatio-temporal variations and source apportionment of water pollution in Danjiangkou Reservoir Basin. *Central China Water, 7*, 2591–2611.

[40] Li, Y., Xu, L., & Li, S. (2009). Water quality analysis of the Songhua River Basin using multivariate techniques. *Journal of Water Resource Protection, 1*, 110

[41] Huang, W. J., Chen, W. Y., Chuang, Y. H., Lin, Y. H., & Chen, H. W. (2014). Biological toxicity of groundwater in a seashore area: Causal analysis and its spatial pollutant pattern. *Chemosphere, 100*, 8–15.

[42] Paatero, P., & Tapper, U. (1994). Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics, 5*, 111–126.

[43] Reff, A., Eberly, S. I., & Bhave, P. V. (2007). Receptor modeling of ambient particulate matter data using positive matrix factorization: Review of existing methods. *Journal of the Air & Waste Management*

Association, 57, 146–154.

[44] USEPA. (2012). *Integrated Risk Information System (IRIS).* National Center for Environmental Assessment, Office of Research and Development, Washington, DC.

[45] Ali, M. M., Ali, M. L., Islam, M. S., & Rahman, M. Z. (2016). Preliminary assessment of heavy metals in water and sediment of Karnaphuli River, Bangladesh. *Environmental Nanotechnology, Monitoring & Management, 5*, 27–35.

[46] Banerjee, S., Kumar, A., Maiti, S. K., & Chowdhury, A. (2016). Seasonal variation in heavy metal contaminations in water and sediments of Jamshedpur stretch of Subarnarekha River, India. *Environmental Earth Sciences, 75*, 265.

[47] Bilgin, A., & Konanç, M. U. (2016). Evaluation of surface water quality and heavy metal pollution of Coruh River Basin (Turkey) by multivariate statistical methods. *Environmental Earth Sciences, 75*, 1029.

[48] Khound, N. J., & Bhattacharyya, K. G. (2017). Multivariate statistical evaluation of heavy metals in the surface water sources of Jia Bharali River Basin, North Brahmaputra Plain, India. *Applied Water Science, 7*, 2577–2586.

[49] Kuang, C., Li, Q., Zhang, Y., Wang, H., & Liu, Z. (2016). Assessment of heavy metal contamination in water body and riverbed sediments of the Yanghe River in the Bohai Sea, China. *Environmental Earth Sciences, 75*, 1105.

[50] Wang, J., Liu, G., Liu, H., & Lam, P. K. (2017). Multivariate statistical evaluation of dissolved trace elements and a water quality assessment in the middle reaches of Huaihe River, Anhui, China. *Science of the Total Environment, 583*, 421–431.

[51] Helena, B., Pardo, R., Vega, M., Barrado, E., Fernandez, J. M., & Fernandez, L. (2000). Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis. *Water Research, 34*, 807–816.

[52] Singh, K. P., Malik, A., Mohan, D., & Sinha, S. (2004). Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)—A case study. *Water Research, 38*(18), 3980–3992.

[53] Huang, J. L., Ho, M. H., & Du, P. F. (2011). Assessment of temporal and spatial variation of coastal water quality and source identification along Macau Peninsula. *Stochastic Environmental Research and Risk Assessment, 25*(3), 353–361.

[54] Zhou, F., Huang, G. H., Guo, H., Zhang, W., & Hao, Z. (2007). Spatio-temporal patterns and source apportionment of coastal water pollution in eastern Hong Kong. *Water Research, 41*(15), 3429–3439.

[55] Carrer, S., & Leardi, R. (2006). Characterizing the pollution produced by an industrial area—Chemometric methods applied to the Lagoon of Venice. *Science of the Total Environment, 370*, 99–116.

[56] Hoinaski, L., Franco, D., Stuetz, R., Sivret, E., & de Melo Lisboa, H. (2013). Investigation of PM10 sources in Santa Catarina, Brazil through graphical interpretation analysis combined with receptor modelling. *Environmental Technology, 34*, 2453–2463.

[57] Henry, R. C., & Christensen, E. R. (2010). Selecting an appropriate multivariate source apportionment model result. *Environmental Science & Technology, 44*(7), 2474–2481.

[58] Hopke, P. K. (2010). The application of receptor modeling to air quality data. *Pollution Atmosphérique.*

[59] Anttila, P., Paatero, P., Tapper, U., & Järvinen, O. (1995). Source identification of bulk wet deposition in Finland by positive matrix factorization. *Atmospheric Environment, 29*, 1705–1718.

[60] Li, H., Hopke, P. K., Liu, X., Du, X., & Li, F. (2015). Application of positive matrix factorization to source apportionment of surface water quality of the Daliao River basin, northeast China. Environmental Monitoring and Assessment, 187, 1–12.

[61] Soonthornnonda, P., & Christensen, E. R. (2008). Source apportionment of pollutants and flows of combined sewer wastewater. *Water Research, 42*, 1989–1998.

[62] Fortner, E., Onasch, T., Canagaratna, M., Williams, L. R., Lee, T., Jayne, J., & Worsnop, D. (2018). Examining the chemical composition of black carbon particles from biomass burning with SAMS. *Journal of Aerosol Science, 120*, 12–21.

[63] Mohr, C., DeCarlo, P. F., Heringa, M. F., Chirico, R., Slowik, J. G., Richter, R., Reche, C., Alastuey, A., Querol, X., Seco, R., Peñuelas, J., Jiménez, J. L., Crippa, M., Zimmermann, R., Baltensperger, U., & Prévôt, A. S. H. (2012). Identification and quantification of organic aerosol from cooking and other sources in Barcelona using aerosol mass spectrometer data. *Atmospheric Chemistry and Physics, 12*, 1649–1665.

[64] Kuang, B. Y., Lin, P., Huang, X. H. H., & Yu, J. Z. (2015). Sources of humic-like substances in the Pearl River Delta, China: Positive matrix factorization analysis of PM2.5 major components and source markers. *Atmospheric Chemistry and Physics, 15*, 1995–2008.

[65] Sowlat, M. H., Hasheminassab, S., & Sioutas, C. (2016). Source apportionment of ambient particle number concentrations in central Los Angeles using positive matrix factorization (PMF). *Atmospheric Chemistry and Physics, 16*, 4849–4866.

[66] Visser, S., Slowik, J. G., Furger, M., Zotter, P., Bukowiecki, N., Canonaco, F., Flechsig, U., Appel, K., Green, D. C., Tremper, A. H., Young, D. E., Williams, P. I., Allan, J. D., Coe, H., Williams, L. R., Mohr, C., Xu, L., Ng, N. L., Nemitz, E., Barlow, J. F., Halios, C. H., Fleming, Z. L., Baltensperger, U., & Prévôt, A. S. H. (2015). Advanced source apportionment of size-resolved trace elements at multiple sites in London during winter. *Atmospheric Chemistry and Physics, 15*, 11291–11309.

[67] Yan, C., Nie, W., Äijälä, M., Rissanen, M. P., Canagaratna, M. R., Massoli, P., Junninen, H., Jokinen, T., Sarnela, N., Häme, S. A. K., Schobesberger, S., Canonaco, F., Yao, L., Prévôt, A. S. H., Petäjä, T., Kulmala, M., Sipilä, M., Worsnop, D. R., & Ehn, M. (2016). Source characterization of highly oxidized multifunctional compounds in a boreal forest environment using positive matrix factorization. *Atmospheric Chemistry and Physics, 16*, 12715–12731.

[68] Comero, S., Servida, D., & De Capitani, L., & Gawlik, B. M. (2012). Geochemical characterization of an abandoned mine site: A combined positive matrix factorization and GIS approach compared with principal component analysis. *Journal of Geochemical Exploration, 118*, 30–37.

[69] González-Macías, C., Sánchez-Reyna, G., Salazar-Coria, L., & Schifter, I. (2014). Application of the positive matrix factorization approach to identify heavy metal sources in sediments: A case study on the Mexican Pacific Coast. *Environmental Monitoring and Assessment, 186*, 307–324.

[70] Li, H., Hopke, P. K., Liu, X., Du, X., & Li, F. (2015). Application of positive matrix factorization to source apportionment of surface water quality of the Daliao River basin, northeast China. *Environmental Monitoring and Assessment, 187*, 80.

[71] Haji Gholizadeh, M., Melesse, A. M., & Reddi, L. (2016). Water quality assessment and apportionment of pollution sources using APCS-MLR and PMF receptor modeling techniques in three major rivers of South Florida. *Science of the Total Environment, 566–567*, 1552–1567.

[72] Zanotti, C., Rotiroti, M., Fumagalli, L., Stefania, G. A., Canonaco, F., Stefenelli, G., Prévôt, A. S. H., Leoni, B., & Bonomi, T. (2019). Groundwater and surface water quality characterization through positive matrix factorization combined with GIS approach. *Water Research, 159*, 122–134.

[73] Jiang, J. U., Khan, A., Shi, B., Tang, S., & Khan, J. (2019). Application of positive matrix factorization to identify potential sources of water quality deterioration of Huaihe River, China. *Applied Water Science, 9*, 63.

[74] Salim, I., Sajjad, R. U., Paule-Mercado, M. C., Memon, S. A., Lee, B. Y., Sukhbaatar, C., & Lee, C. H. (2019). Comparison of two receptor models PCA-MLR and PMF for source identification and apportionment of pollution carried by runoff from catchment and sub-watershed areas with mixed land cover in South Korea. *Science of the Total Environment, 663*, 764–775.

[75] Sun, X., Wang, H., Guo, Z., Lu, P., Song, F., Liu, L., Liu, J., Rose, N. L., & Wang, F. (2020). Positive matrix factorization on source apportionment for typical pollutants in different environmental media: A review. *Environmental Science: Processes & Impacts, 22*(3), 239–255.