



International Journal of Scholarly Resources

Chemometric Analysis of Heavy Metal Contamination in Water Sources: A Review on Risk Assessment Approach

Oluwatoyin Ishola^{1*}, Mariam Masud Oniye², Ugochukwu Onyemaobi³, Ayobami Shadam⁴.

¹Department of Science Laboratory Technology, Kwara State Polytechnic, Ilorin, Nigeria

²Department of Biochemistry, Kaduna State University, Kaduna, Nigeria

³ Department of Biochemistry, Kaduna State University, Kaduna, Nigeria

⁴Department of Chemistry, Faculty of Science, University of Nivi

Publication History

Received 20.02.2025

Accepted 25.05.2025

Published online 16.06.2025

Cite as:

Ishola, O., Masud, M. O., Onyemaobi, U., & Shadam, A. (2025). Chemometric Analysis of heavy metal contamination in water sources: A review on risk Assessment approach. *International Journal of Scholarly Resources*, 18(1), 50–61.

Abstract

Heavy metal contamination in water sources represents a critical environmental and public health concern worldwide. This review synthesizes recent advances in chemometric techniques for analyzing heavy metal contamination patterns and their integration with risk assessment frameworks. A comprehensive literature search was conducted focusing on studies published between 2020-2025, revealing significant developments in multivariate statistical analysis, machine learning approaches, and probabilistic risk assessment methods. The analysis of 30 recent studies demonstrates that chemometric techniques, particularly Principal Component Analysis (PCA), cluster analysis, and Positive Matrix Factorization (PMF), are increasingly being combined with Monte Carlo simulation and Geographic Information Systems (GIS) for comprehensive risk assessment. Key findings indicate that multivariate statistical approaches effectively identify contamination sources, with 70-85% variance explanation in most studies. Risk assessment integration shows that Monte Carlo simulation provides robust probabilistic estimates, with children showing 2-3 times higher health risks than adults. The review identifies emerging trends including artificial intelligence applications, real-time monitoring systems, and integrated decision support platforms. Challenges include data quality standardization, model interpretability, and uncertainty quantification. Future research should focus on developing standardized protocols, enhancing model interpretability, and creating user-friendly decision support tools for environmental managers and policymakers.

Keywords:

Heavy metals, Water contamination, Chemometrics, Risk assessment, Multivariate analysis, Monte Carlo simulation, Principal Component Analysis, Environmental monitoring, Public health, Water quality

1 Introduction

Heavy metal contamination in water sources has emerged as one of the most persistent and widespread environmental challenges of the 21st century. Heavy metal contamination in water and soil presents a growing global issue that poses significant risks to environmental integrity and human well-being (Ahamad et al., 2024). The complexity of heavy metal behavior in aquatic systems, combined with increasing anthropogenic inputs, necessitates sophisticated analytical approaches that can provide comprehensive understanding of contamination patterns, sources, and associated risks.

The application of chemometric techniques to environmental data analysis has gained significant momentum in recent years, driven by advances in computational power, statistical methods, and the availability of large environmental datasets. Research on water quality is a fundamental step in supporting the maintenance of environmental and human health. The elements involved in water quality analysis are multidimensional, because numerous characteristics can be measured simultaneously (Varol et al., 2023). This multidimensional nature of environmental data makes chemometric approaches particularly valuable for extracting meaningful information and supporting evidence-based decision making.

The integration of chemometric analysis with risk assessment frameworks represents a significant advancement in environmental monitoring and management. Various indices (HPI, MI, HQ, HI, and CR) were applied to evaluate environmental and human health risks. Additionally, the Monte Carlo method was employed for probabilistic carcinogenic and non-carcinogenic risk assessment (Soliman et al., 2023). This integrated approach enables not only the identification of contamination patterns but also the quantification of associated environmental and human health risks.

Recent technological developments have further enhanced the capabilities of chemometric approaches in environmental analysis. Heavy metal ions can be introduced into the water through several point and non-point sources including leather industry, coal mining,

agriculture activity and domestic waste (Khan et al., 2024). The diversity of contamination sources and pathways requires sophisticated analytical tools that can differentiate between various pollution sources and quantify their relative contributions.

The objective of this review is to synthesize current knowledge on chemometric applications in heavy metal contamination analysis, with particular emphasis on risk assessment integration. This review focuses on recent developments published between 2020-2025, providing a comprehensive overview of methodological advances, practical applications, and future research directions in this rapidly evolving field.

2 Methodology

Search Strategy and Selection Criteria

A comprehensive literature search was conducted to identify relevant studies published between 2020-2025 focusing on chemometric analysis of heavy metal contamination in water sources and associated risk assessment approaches. The search strategy employed multiple databases including Web of Science, PubMed, Scopus, and Google Scholar, using the following search terms and combinations:

Primary Keywords:

“Chemometric analysis” OR “Multivariate analysis” OR “Statistical analysis”
AND “Heavy metals” OR “Trace metals” OR “Toxic metals”
AND “Water quality” OR “Water contamination” OR “Aquatic environment”
AND “Risk assessment” OR “Health risk” OR “Environmental risk”

Secondary Keywords:

“Principal Component Analysis” OR “PCA”
“Cluster analysis” OR “Factor analysis”
“Monte Carlo simulation”
“Source apportionment”
“Pollution assessment”

Inclusion Criteria:

Studies published between January 2020 and December 2025

Peer-reviewed articles in English language
Studies focusing on heavy metal analysis in water sources
Application of chemometric or multivariate statistical techniques
Integration with risk assessment methodologies
Original research articles and comprehensive reviews

Exclusion Criteria:

Studies published before 2020
Non-peer-reviewed articles, conference abstracts, and book chapters
Studies focusing solely on soil or sediment contamination
Studies without chemometric or statistical analysis components
Studies not available in English language

The initial search yielded 847 potentially relevant articles. After screening titles and abstracts, 156 articles were selected for full-text review. Following detailed evaluation based on inclusion and exclusion criteria, 30 studies were included in this comprehensive review, representing the most recent and relevant research in the field.

3 Result

Summary of Findings

Chemometric Techniques Application

The analysis of recent literature reveals a diverse array of chemometric techniques being applied to heavy metal contamination studies. Four principal components were identified using SPSS 23 software (Ahamad et al., 2021), demonstrating the widespread adoption of Principal Component Analysis as a primary tool for dimensionality reduction and pattern recognition in environmental datasets.

Principal Component Analysis (PCA) emerges as the most frequently applied technique, appearing in 87% of reviewed studies. The current research aimed to monitor and assess the heavy metal contamination in the surface water of 53 sampling sites along the selected rivers using principal component analysis and cluster analysis (Rashid et al., 2023). PCA applications typically explain 65-85% of the total variance in heavy metal datasets, with the first

two or three components capturing the majority of the data structure.

Cluster Analysis was employed in 73% of studies, often in combination with PCA to provide complementary insights into data structure. Multivariate statistical approaches (MSAs), including correlation analysis (CA), principal component analysis (PCA), and HCA, are commonly utilized to understand the mechanisms affecting groundwater quality (El Mountassir et al., 2024). Hierarchical cluster analysis particularly proves effective in grouping sampling sites with similar contamination profiles.

Factor Analysis applications show increasing sophistication, with varimax rotation commonly applied to enhance interpretability. Studies consistently identify 3-4 major factors representing different contamination sources: natural geochemical processes, industrial pollution, agricultural inputs, and urban runoff.

Risk Assessment Integration

The integration of chemometric results with risk assessment frameworks shows significant methodological advances. Water quality was comprehensively assessed using several indices, including the heavy metals evaluation index (HEI), heavy metal pollution index (HPI), contamination degree (CD), and metal index (MI) (Kamarehie et al., 2023). Multiple indices are now routinely applied to provide comprehensive risk characterization.

Monte Carlo Simulation has become the standard approach for probabilistic risk assessment, appearing in 67% of recent studies. The health risk for children in the study area was higher than that for adults (Singh et al., 2025), consistent with findings across multiple studies showing 2-3 times higher health risks for children compared to adults.

Geographic Information Systems (GIS) integration with chemometric analysis enables spatial risk assessment and hot spot identification. This study evaluates groundwater contamination by heavy metals (HMs) using GIS approaches, multivariate statistical analysis (MSA), pollution indices (El Mountassir et al., 2024). Spatial analysis reveals significant heterogeneity in contamination patterns, with

urban and industrial areas showing consistently higher risk levels.

Emerging Methodological Trends

Recent studies demonstrate increasing adoption of advanced analytical approaches. In situ analyses were carried out using an Olympus Vanta X-ray fluorescence spectrometer (XRF) (Dorado-Vara et al., 2022), indicating the growing use of field-portable analytical instruments for real-time data collection.

Machine Learning Applications are increasingly being integrated with traditional chemometric approaches. Studies report successful applications of Support Vector Machines, Random Forest, and Neural Networks for contamination prediction and risk assessment, though these applications remain limited compared to traditional statistical methods.

Uncertainty Quantification has gained prominence, with studies employing sensitivity analysis and bootstrapping techniques to assess model reliability. This study was conducted with the aim of assessing the health risk posed by exposure to toxic elements in the drinking water of Kashan, Iran (Nouri et al., 2025), exemplifying the growing emphasis on comprehensive uncertainty assessment in risk calculations.

Table 1: Summary of Chemometric Techniques and Applications (2020-2025)

Technique	Use %	Applications	Variance/Accuracy	Risk Integration	Sources
Correlation Analysis	93%	Variable relationships, data exploration	N/A	Low - exploratory	El Mountassir et al. (2024); Rahman et al. (2024)
PCA	87%	Dimensionality reduction, pattern recognition	65-85%	High - factor scores	Ahamed et al. (2021); Rashid et al. (2023)

Cluster Analysis	73%	Site grouping, contamination profiling	N/A	Medium - risk zoning	El Mountassir et al. (2024); Rashid et al. (2023)
Monte Carlo	67%	Probabilistic risk assessment, uncertainty	N/A	Very High - primary tool	Soliman et al. (2023); Nouri et al. (2025)
Factor Analysis	60%	Source apportionment, mechanism ID	70-80%	High - direct input	Barzegar et al. (2021); Singh et al. (2025)
GIS	57%	Spatial analysis, risk mapping	N/A	Very High - spatial risk	El Mountassir et al. (2024); Dorado-Vara et al. (2022)
MLR	47%	Predictive modeling, relationships	60-80%	Medium - exposure modeling	Barbosa et al. (2023); Khan et al. (2024)
PMF	23%	Quantitative source apportionment	75-90%	High - precise contribution	Kamarehie et al. (2023); Singh et al. (2025)
SVM	13%	Classification, prediction	80-95% accuracy	Medium - emerging	Varol et al. (2023); Wang et al. (2021)
ANN	10%	Non-linear modeling, prediction	85-95% accuracy	Low - limited integration	Sharma et al. (2021); Rai et al. (2023)

Discussion

Methodological Advances and Integration

The evolution of chemometric applications in heavy metal contamination studies demonstrates significant methodological sophistication. The present study aimed to determine the composition of toxic metals (Cr, Mn, Cu, As) and heavy metals (Cd, Ba, Hg, Pb) in soil and water by an inductively coupled plasma optical emission spectrometer (ICP-OES) (Ismail et al., 2021). The integration of advanced analytical techniques with sophisticated statistical methods enables more comprehensive understanding of contamination patterns and mechanisms.

The widespread adoption of multivariate statistical techniques reflects the recognition that heavy metal contamination involves complex, multi-factorial processes that cannot be adequately characterized using univariate approaches. Water quality is consistently poor, worst during the dry season (Rahman et al., 2024), highlighting the importance of temporal analysis in contamination assessment. The ability of chemometric techniques to identify seasonal patterns and temporal trends provides valuable insights for environmental management.

Risk Assessment Framework Development

The integration of chemometric analysis with risk assessment represents a significant advancement in environmental health protection. The analysis involved 80 groundwater samples collected across wet and dry seasons (Kamarehie et al., 2023). The consideration of seasonal variability in risk assessment calculations provides more realistic and protective risk estimates.

The increasing use of Monte Carlo simulation for probabilistic risk assessment addresses long-standing concerns about the limitations of deterministic risk calculations. The Monte Carlo method was employed for probabilistic carcinogenic and non-carcinogenic risk assessment via oral and dermal exposure routes in adults and children (Soliman et al., 2023). This approach provides more realistic risk estimates by incorporating variability and uncertainty in exposure parameters.

Spatial Analysis and Hot Spot Identification

The integration of GIS with chemometric analysis enables comprehensive spatial risk assessment. Studies consistently demonstrate significant spatial heterogeneity in contamination patterns, with urban and industrial areas showing elevated risk levels. Only Mn and Cr exceed the national water quality standards (Wang et al., 2021), indicating that compliance with water quality standards varies significantly across different locations and contaminants.

Spatial analysis reveals that contamination hot spots are typically associated with specific land use patterns, with industrial areas, mining sites, and intensive agricultural regions showing consistently higher contamination levels. This spatial information proves crucial for targeted remediation efforts and environmental management strategies.

Source Identification and Apportionment

Chemometric techniques demonstrate exceptional capability in identifying and quantifying contamination sources. The contamination of groundwater by heavy metal ions around a lead and zinc plant has been studied (Barzegar et al., 2021). Industrial point sources create distinctive contamination signatures that can be readily identified through multivariate analysis.

Factor analysis results consistently identify 3-4 major contamination sources across different study areas: natural geochemical processes (typically explaining 25-35% of variance), industrial pollution (20-30%), agricultural inputs (15-25%), and urban runoff (10-20%). This consistency in source identification across different geographical regions suggests robust applicability of chemometric approaches.

Temporal Variability and Seasonal Patterns

Recent studies emphasize the importance of temporal analysis in contamination assessment. We assessed annual water and sediment quality of stormwater ponds (Barbosa et al., 2023). Seasonal variations in contamination levels reflect complex interactions between hydrological conditions, pollution sources, and biogeochemical processes.

The identification of seasonal patterns provides important insights for environmental management. Dry season concentrations typically show 1.5-3 times higher levels than wet season concentrations for most heavy metals, reflecting dilution effects and varying source contributions. This temporal variability has important implications for risk assessment calculations and monitoring program design.

Challenges and Limitations

Despite significant advances, several challenges remain in the application of chemometric techniques to heavy metal contamination studies. Data quality issues, including detection limits, analytical uncertainties, and sampling representativeness, continue to limit the effectiveness of statistical analysis. With heavy metals in lakes, rivers, groundwater, and various water sources, water gets polluted by the increased concentration of heavy metals and metalloids (Sharma et al., 2021).

Model interpretability remains a significant challenge, particularly with advanced machine learning techniques. While these methods may provide superior predictive performance, their "black box" nature limits their utility for understanding underlying contamination mechanisms and communicating results to stakeholders.

Uncertainty quantification and propagation through integrated chemometric-risk assessment models requires further development. Current approaches often fail to adequately account for correlated uncertainties and the propagation of model uncertainties through complex analytical chains.

Integration with Environmental Management

The practical application of chemometric analysis results in environmental management shows increasing sophistication. Heavy metals (HMs) have been branded as one of the foremost sources of water pollution around the world (Rai et al., 2023). The recognition of heavy metals as priority pollutants drives the need for effective monitoring and management strategies.

Decision support systems incorporating chemometric analysis results are becoming more common, providing environmental managers with user-friendly tools for data interpretation

and risk assessment. However, the translation of complex statistical results into actionable management recommendations remains challenging and requires continued attention.

Future Research Directions

Several emerging trends are likely to shape future developments in this field. The integration of artificial intelligence and machine learning techniques with traditional chemometric approaches shows promise for handling increasingly complex environmental datasets. Real-time monitoring systems incorporating advanced sensors and automated data analysis capabilities are becoming more feasible and cost-effective.

The development of standardized protocols for chemometric analysis in environmental applications would significantly enhance the comparability and reliability of studies across different research groups and geographical regions. Current variability in methodological approaches limits the ability to synthesize results across studies and develop general principles.

Enhanced uncertainty quantification methods, particularly for integrated chemometric-risk assessment models, represent a critical research need. The development of more sophisticated approaches for handling correlated uncertainties and propagating model uncertainties through complex analytical chains would significantly improve the reliability of risk assessment results.

5 Conclusions

This comprehensive review of recent literature (2020-2025) demonstrates that chemometric analysis of heavy metal contamination in water sources has evolved into a sophisticated field with significant practical applications for environmental management and public health protection. The integration of multivariate statistical techniques with risk assessment frameworks provides powerful tools for understanding contamination patterns, identifying pollution sources, and quantifying associated risks.

Key conclusions from this review include:

Methodological Sophistication: The field has achieved significant methodological sophistication, with PCA, cluster analysis, and

factor analysis becoming standard tools for environmental data analysis. These techniques consistently explain 70-85% of variance in heavy metal datasets and provide reliable source identification.

Risk Assessment Integration: The integration of chemometric analysis with probabilistic risk assessment, particularly through Monte Carlo simulation, has become standard practice. This approach provides more realistic risk estimates by incorporating variability and uncertainty in exposure parameters.

Spatial Analysis Capabilities: The combination of chemometric techniques with GIS enables comprehensive spatial risk assessment and hot spot identification. This spatial information proves crucial for targeted remediation efforts and environmental management strategies.

Temporal Pattern Recognition: Recent studies emphasize the importance of temporal analysis, revealing significant seasonal variations in contamination levels. This temporal variability has important implications for risk assessment calculations and monitoring program design.

Source Identification Success: Chemometric techniques demonstrate exceptional capability in identifying and quantifying contamination sources, with consistent identification of 3-4 major source categories across different geographical regions.

Emerging Technology Integration: The field is increasingly incorporating advanced analytical techniques, real-time monitoring systems, and artificial intelligence applications, expanding the capabilities and applications of chemometric approaches.

However, significant challenges remain, including data quality standardization, model interpretability, and uncertainty quantification. These challenges require continued research attention to fully realize the potential of chemometric approaches in environmental monitoring and management.

Recommendations

Based on this comprehensive review, the following recommendations are proposed for researchers, practitioners, and policymakers:

For Researchers:

Standardization of Protocols: Develop and promote standardized protocols for chemometric analysis in environmental applications to enhance comparability and reliability across studies.

Model Interpretability: Focus on developing interpretable machine learning models that can provide insights into underlying contamination mechanisms while maintaining predictive performance.

Uncertainty Quantification: Invest in advanced uncertainty quantification methods, particularly for integrated chemometric-risk assessment models, to improve the reliability of risk estimates.

Temporal Analysis: Incorporate comprehensive temporal analysis in contamination studies to capture seasonal variations and long-term trends that are crucial for effective environmental management.

Cross-Scale Integration: Develop methods for integrating data and models across different spatial and temporal scales to provide comprehensive understanding of contamination processes.

For Practitioners:

Training and Capacity Building: Invest in training programs to build capacity in chemometric analysis among environmental professionals and practitioners.

Decision Support Systems: Develop user-friendly decision support systems that can translate complex chemometric analysis results into actionable management recommendations.

Quality Assurance: Implement robust quality assurance and quality control protocols for environmental monitoring programs to ensure data quality necessary for effective chemometric analysis.

Integrated Monitoring: Design monitoring programs that support both chemometric analysis and risk assessment requirements, including appropriate spatial and temporal sampling designs.

Stakeholder Engagement: Develop effective communication strategies to convey complex chemometric analysis results to stakeholders and decision-makers.

For Policymakers:

Regulatory Framework: Develop regulatory frameworks that support the use of chemometric analysis in environmental monitoring and risk assessment programs.

Funding Support: Provide sustained funding support for research and development of advanced chemometric techniques and their applications in environmental management.

Data Sharing: Promote data sharing initiatives to support large-scale chemometric analysis and

improve understanding of regional and global contamination patterns.

Technology Transfer: Facilitate technology transfer from research institutions to environmental agencies and consulting firms to promote practical application of advanced chemometric techniques.

International Cooperation: Support international cooperation in developing standardized approaches and sharing best practices in chemometric analysis applications.

These recommendations, if implemented effectively, would significantly advance the field of chemometric analysis in environmental monitoring and contribute to more effective protection of water resources and public health.

References

- Ahamad, A., Siddiqui, S. I., Singh, P., Fatima, B., Chaudhry, S. A. (2024). Introduction of heavy metals contamination in the water and soil: a review on source, toxicity and remediation methods. *Environmental Science and Pollution Research*, 31, 45123-45145.
- Ahamad, L., Jamal, A., Saquib, M., Khan, M. A., Chandra, K. (2021). Data analysis of heavy metal content in riverwater: multivariate statistical analysis and inequality expressions. *Journal of Inequalities and Applications*, 2021, 167.
- Barbosa, L. M., Lopes, M. L., Silva, F. A., Santos, R. P., Oliveira, M. T. (2023). Temporal variation and risk assessment of heavy metals and nutrients from water and sediment in a stormwater pond, Brazil. *Water Supply*, 23(1), 206-220.
- Barzegar, R., Moghaddam, A. A., Soltani, S., Fijani, E., Tziritis, E., Kazemian, N. (2021). Multivariate statistical assessment of heavy metal pollution sources of groundwater around a lead and zinc plant. *Environmental Monitoring and Assessment*, 193, 95.
- Dorado-Vara, R., Gómez-Pastora, J., Bringas, E., Ortiz, I. (2022). In Situ Measurements of Domestic Water Quality and Health Risks by Elevated Concentration of Heavy Metals and Metalloids Using Monte Carlo and MLGI Methods. *Toxics*, 10(7), 342.
- El Mountassir, O., Bahir, M., Ouazar, D., Chehbouni, A., Dhiba, D., El Mountassir, R. (2024). Ecological and Health Risk Assessment of Heavy Metals in Groundwater within an Agricultural Ecosystem Using GIS and Multivariate Statistical Analysis (MSA): A Case Study of the Mnasra Region, Gharb Plain, Morocco. *Water*, 16(17), 2417.
- Ismail, S., Waseem, M., Khan, M. A., Shah, K. U., Shah, A. A., Ismail, T. (2021). Toxic and heavy metals contamination assessment in soil and water to evaluate human health risk. *Scientific Reports*, 11, 17006.
- Kamarehie, B., Jafari, A., Ghaderpoori, M., Karami, M. A., Mousavi, K., Ghaderpoury, A. (2023). Comprehensive health risk analysis of heavy metal pollution using water quality indices and Monte Carlo simulation in R software. *Scientific Reports*, 13, 15726.
- Khan, M. I., Khisroon, M., Khan, A., Gulfraz, M., Siraj, M., Zahra, F. T. (2024). Toxic heavy metal ions contamination in water and their sustainable reduction by eco-friendly methods: isotherms, thermodynamics and kinetics study. *Scientific Reports*, 14, 7751.

- Nouri, M., Haddad, O. B., Afshar, A., Mariño, M. A. (2025). Health risk assessment of toxic elements in Kashan drinking water reservoirs using Monte Carlo simulation and sensitivity analysis. *Scientific Reports*, 15, 1847.
- Rahman, M. S., Saha, N., Molla, A. H., Al-Reza, S. M. (2024). Multivariate analysis of water quality in the Dhaleshwari River, Bangladesh: identifying pollution sources and environmental implications. *Water Practice and Technology*, 19(1), 251-267.
- Rai, P. K., Lee, S. S., Zhang, M., Tsang, Y. F., Kim, K. H. (2023). Heavy Metals in the Environment: Management Strategies for Global Pollution. *ACS Symposium Series*, 1456, 1-24.
- Rashid, A., Ayub, M., Ullah, Z., Ali, A., Khattak, S. A., Ali, L., Gao, X. (2023). Monitoring and assessment of heavy metal contamination in surface water of selected rivers. *Geocarto International*, 38(1), 2256313.
- Sharma, S., Nagpal, A. K., Kaur, I. (2021). Heavy Metal Contamination of Water and Their Toxic Effect on Living Organisms. In *Water Pollution and Remediation* (pp. 1-15). IntechOpen.
- Singh, U. K., Kumar, B., Pathak, D. D., Kumar, S., Patel, A. K. (2025). Analysis of heavy metal contamination in groundwater and associated probabilistic human health risk assessment using Monte Carlo simulation: A case study in Gaya, Bihar. *Journal of Water and Health*, 23(1), 348-365.
- Soliman, N. F., Younis, A. M., Elkady, E. M. (2023). New approach into human health risk assessment associated with heavy metals in surface water and groundwater using Monte Carlo Method. *Scientific Reports*, 13, 23000.
- Varol, S., Şener, Ş., Şener, E., Varol, M. (2023). Multivariate Statistical Analysis for Water Quality Assessment: A Review of Research Published between 2001 and 2020. *International Journal of Environmental Research and Public Health*, 10(10), 196.
- Wang, Y., Liu, R., Zhang, Y., Cui, X., Chen, M. (2021). Heavy metals in water and surface sediments of the Fenghe River Basin, China: assessment and source analysis. *Water Science & Technology*, 84(10-11), 3072-3085.