DOI:10.5937/IIZS25300B

UNSUPERVISED MACHINE LEARNING vs. CLASSICAL CLUSTERING: HOTSPOT IDENTIFICATION OF PBDEs IN DANUBE RIVER SEDIMENTS

Maja Brborić[1], Branka Nakomčić Smaragdakis[1], Sonja Dmitrašinović[1], Borivoje Stepanov[2], Epari Ritesh Patro[3], Maja Turk Sekulić[1]
[1]University of Novi Sad, Faculty of Technical Sciences, Department of Environmental Engineering and Occupational Safety and Health, Novi Sad, Serbia
[2]University of Novi Sad, Faculty of Technical Sciences, Department of Energy and Process Technique, Novi Sad, Serbia
[3]Water, Energy and Environmental Engineering Research Unit, University of Oulu, Oulu FI90014, Finland
e-mail: majabrboric@uns.ac.rs

Abstract: This study investigates the concentrations of ten PBDE congeners in Danube River bottom sediments and identifies contamination hotspots using hierarchical cluster analysis (HCA) and k-means clustering. Σ10PBDE ranged from 0.52 to 31.21 μg/kg d.w., with BDE-209 dominating. HCA grouped sites according to congener profiles and overall burden, while k-means, based on a neural network learning principle, provided enhanced resolution of high-risk sites. Both approaches identified Neštin, D6 and D10 as hotspots. The combined application of classical and machine-learning clustering is recommended for improved prioritisation of monitoring and management efforts.
Key words: PBDEs, sediment, Danube, water pollution

INTRODUCTION

River sediments are key environmental compartments that act as both sinks and potential secondary sources of hydrophobic contaminants. Due to their fine-grained structure and high organic carbon content, sediments accumulate lipophilic pollutants such as polycyclic aromatic hydrocarbons (PAHs), polychlorinated biphenyls (PCBs), and brominated flame retardants (BFRs) [1,2]. Once deposited, these substances can persist for years, undergoing remobilisation during hydrological events, dredging, or resuspension, thereby reintroducing contaminants into the water column and food web [3].
Polybrominated diphenyl ethers (PBDEs), widely used as flame retardants in plastics, textiles and electronics, are of particular concern because of their persistence, bioaccumulation potential, and toxicity. Laboratory and field studies have linked PBDE exposure to endocrine disruption, neurodevelopmental deficits, and carcinogenic effects [4,5]. In aquatic ecosystems, PBDEs can accumulate in sediments, where they pose risks to benthic organisms and enter the trophic chain via bioaccumulation and biomagnification [6]. The Danube River, as Europe's second largest river, flows through multiple countries and receives diverse anthropogenic inputs, making it a crucial hotspot for transboundary pollution studies.
The analysis of PBDEs in sediments is not only important for establishing contamination levels but also for identifying critical sites ("hotspots") that require focused monitoring and management. Statistical and machine-learning approaches have been increasingly employed to classify sediment quality data and detect hidden structures in complex datasets [7,8]. Hierarchical cluster analysis (HCA) has traditionally been used to group sampling sites based on similarity in contaminant profiles, whereas k-means clustering—rooted in artificial neural networks and unsupervised machine learning—offers a more advanced, iterative optimisation that can reveal subtle boundaries between clusters [9].
This study aims to (i) quantify PBDE concentrations in Serbian Danube sediments, (ii) apply both HCA and k-means to identify contamination hotspots, and (iii) compare the effectiveness of these two approaches. By integrating conventional multivariate statistics with neural

network-based clustering, the research demonstrates the value of combining classical and modern data analysis tools in environmental monitoring.

MATERIAL AND METHODS

Sediment sampling and analysis

Sediment samples (0–10 cm, <63 µm fraction) were collected from ten sites along the Serbian Danube (Fig.1). Composite samples were homogenised, freeze-dried and extracted (Soxhlet, toluene), with clean-up by silica/charcoal columns. PBDE congeners were quantified by GC-MS. Surrogate recoveries were 78–112%, blanks were below LOQ, and precision was within ±15%. Concentrations are reported on a dry-weight basis.

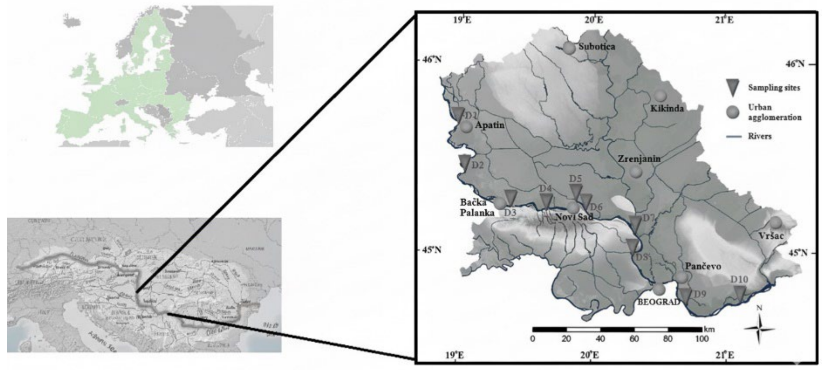

Fig. 1. *Map of the sampling sites*

Statistical analysis

Concentration data (Σ10PBDE and individual congeners) were standardised and processed using SPSS and MATLAB.
Hierarchical cluster analysis (HCA): Ward's linkage with squared Euclidean distances.
k-means clustering: iterative partitioning into 3–5 groups, optimised using silhouette scores. The method follows the learning principle of unsupervised artificial neural networks and represents a fundamental building block of deep learning algorithms [9,10].

RESULTS AND DISCUSSION

Concentration ranges and congener profiles

PBDEs were detected in all sediment samples, with Σ10PBDE concentrations spanning a broad range from 0.52 to 31.21 µg/kg d.w., and a mean value of 8.09 µg/kg. Such variability reflects both spatial heterogeneity in contaminant inputs and differences in sediment properties, including grain size and organic matter content. The congener profile was dominated by BDE-209 (Figure 2), which is consistent with the widespread historical use of technical deca-BDE formulations in plastics, textiles, and electronic housings. BDE-209 is highly hydrophobic and persistent, and its prevalence in sediments indicates both ongoing release from materials still in use and its long-term accumulation in depositional zones.
At several locations, additional congeners made notable contributions. BDE-47 and BDE-99, two of the major constituents of penta-BDE technical mixtures, were detected at elevated proportions in specific sites, suggesting localised inputs from urban or industrial discharges, as well as possible in-situ debromination of higher congeners. The occurrence of BDE-183, typically associated with octa-BDE mixtures, was particularly enriched at site D6, indicating potential inputs from polymer processing or waste-related sources, such as leachates from obsolete electronics and construction materials. In contrast, D10 exhibited a strong dominance

of BDE-209 with limited contributions from lower congeners, pointing to point-source contamination dominated by deca-BDE.
The highest Σ10PBDE concentration was observed in Neštin (D3), where both absolute levels and relative enrichment of mid-brominated congeners suggest upstream transport combined with local sedimentary retention. Taken together, these compositional patterns highlight that technical deca-BDE formulations were the primary sources of PBDEs in the investigated stretch of the Danube, while the detectable presence of tetra- to hepta-BDE congeners reflects secondary inputs from penta- and octa-mixtures, as well as environmental transformation processes such as photolytic or microbial debromination. These findings emphasise that sedimentary PBDE profiles can provide insights not only into contaminant levels but also into the relative importance of historical use, ongoing emissions, and degradation pathways.

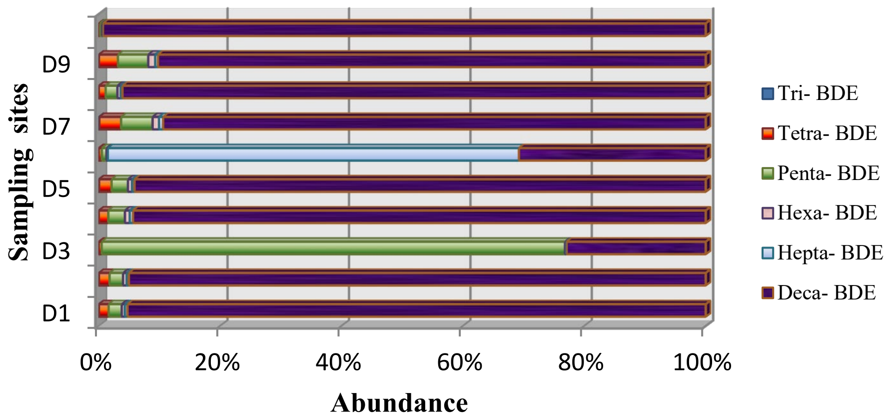

Fig. 2. *Distribution of PBDEs in the riverbed sediment from Danube River*

Hotspot identification by HCA

Hierarchical cluster analysis (HCA) grouped the sampling sites into three major clusters (Fig. 3), providing an intuitive overview of similarity patterns across the Danube sediment dataset. Cluster I comprised sites with low Σ10PBDE contamination, generally below 2 μg/kg, where congener profiles were dominated by trace levels of BDE-209 and negligible contributions of other congeners. These locations can be considered background or reference sites, reflecting either limited anthropogenic influence or depositional environments less prone to contaminant accumulation.
Cluster II included sites with moderate PBDE concentrations (2–10 μg/kg), in which BDE-209 was clearly the predominant congener. This cluster reflects areas where deca-BDE inputs are present but not extreme, likely corresponding to diffuse sources such as runoff from urbanised catchments, abrasion of treated materials, or gradual release from industrial and municipal waste streams. Importantly, the sites in this cluster displayed relatively homogeneous profiles, suggesting a common source signature dominated by technical deca-BDE.
Cluster III contained the identified hotspots: Neštin (D3), D6, and D10. These sites showed markedly higher Σ10PBDE levels, reaching up to 31.21 μg/kg, and exhibited distinct compositional patterns. Neštin (D3) was characterised by both elevated concentrations and the presence of mid-brominated congeners (BDE-47, -99), indicating upstream transport coupled with local depositional processes. Site D6 was enriched in BDE-183, pointing to contributions from octa-BDE mixtures or e-waste-related leachates, while D10 was dominated by BDE-209, consistent with a localised deca-BDE input.
The dendrogram provided a clear hierarchical framework to visualise these relationships, where the progressive branching illustrated how sites with similar contaminant signatures clustered together. This structure is particularly useful for identifying not only extreme hotspots but also intermediate contamination groups, enabling prioritisation of monitoring efforts. By revealing both shared patterns and site-specific anomalies, HCA offered a straightforward but

powerful tool for evaluating the spatial organisation of PBDE contamination in the Danube sediments.

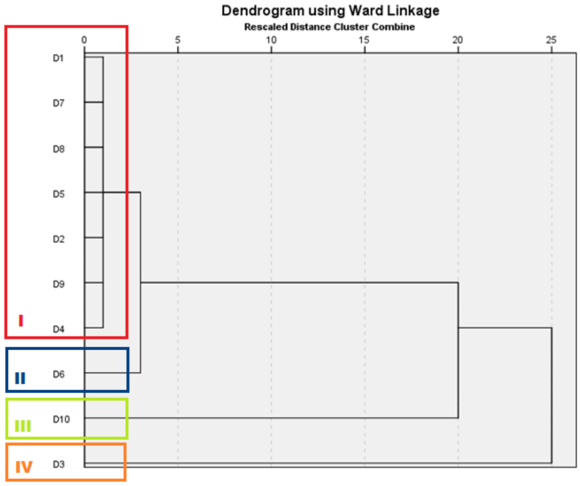

Fig.3. *Cluster dendograms of PBDEs identified in ten tested samples*

Hotspot identification by k-means clustering

The application of k-means clustering produced more nuanced and differentiated groupings compared to HCA, particularly when dealing with sites that exhibited similar Σ10PBDE concentrations but diverged in their congener composition. While HCA provided a broader overview of relationships, k-means, through its iterative partitioning algorithm, succeeded in capturing subtle differences among sites, reflecting its higher sensitivity to internal cluster variability.

Neštin (D3) consistently emerged as the centroid of a high-risk cluster, which confirmed its status as the most contaminated site within the studied river stretch. The elevated Σ10PBDE concentration, combined with the presence of both deca- and penta-BDE related congeners, positioned D3 as a focal point of concern. In contrast, D6 and D10 were identified as distinct outliers, each characterised by unique contaminant profiles. D6 showed a notable enrichment of BDE-183, suggesting inputs from octa-BDE formulations or leachates associated with e-waste disposal, whereas D10 was dominated almost exclusively by BDE-209, pointing to a strong, localised deca-BDE source.

This ability of k-means to highlight intra-cluster variability and distinguish between superficially similar sites illustrates the advantage of neural-network-based clustering over classical hierarchical approaches. Unlike HCA, k-means does not rely on a static tree-like structure but continuously updates group assignments until the optimal partitioning is achieved. This iterative optimisation, rooted in the principles of unsupervised machine learning, ensures that the method can adapt to complex and non-linear patterns in environmental datasets. Furthermore, k-means allows probabilistic interpretation of cluster membership, which is particularly relevant in transitional zones where sites may display characteristics of both moderately contaminated and hotspot groups. Such flexibility provides a richer and more realistic representation of sediment contamination dynamics.

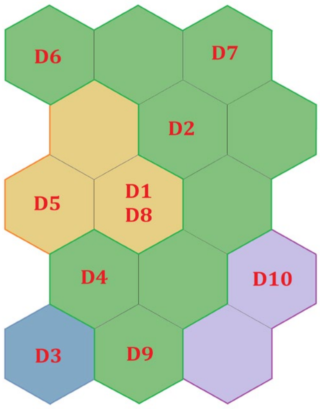

Fig.4. *k-means algorithm for PBDEs*

Comparison of methods

Both HCA and k-means clustering effectively identified contamination hotspots along the Danube River, yet their interpretive strengths and practical applications differ. HCA provided a straightforward, hierarchical visualisation of site similarities, enabling researchers to rapidly identify groups of sites that share common contamination signatures. Its dendrogram structure is simple to interpret and particularly useful in the initial exploratory phase of data analysis, where a broad overview of patterns is required. However, the rigid, stepwise nature of HCA means that once branches are formed, they cannot be re-adjusted, which may lead to oversimplification in cases where subtle variability exists between sites.
K-means clustering, on the other hand, offered a more robust and dynamic approach. Its neural network origin and iterative optimisation process make it well suited for handling larger, multidimensional datasets where complex patterns are present. The method was particularly effective in differentiating sites with similar overall PBDE concentrations but distinct compositional profiles, providing sharper resolution of contamination sources. Its conceptual link to deep learning methodologies underscores its adaptability, as k-means can be integrated with more advanced machine-learning frameworks for predictive modelling of contaminant behaviour.
In practical terms, HCA excels as an exploratory, easy-to-visualise tool, while k-means provides deeper insights and higher resolution in clustering outcomes. When applied together, they complement one another: HCA offers interpretability and simplicity, while k-means adds precision and adaptability. This combined application ensures that both general patterns and finer-scale distinctions are captured, thereby enhancing the reliability of hotspot identification and supporting more informed decision-making in sediment monitoring and management.

CONCLUSIONS

This study confirmed the heterogeneous distribution of PBDEs in Serbian Danube sediments, with concentrations reaching up to 31.21 μg/kg d.w. and BDE-209 consistently dominating the congener profile. Such findings highlight the continuing environmental presence of deca-BDE mixtures, with additional contributions from lower-brominated congeners pointing to both historical use and possible debromination processes.
The application of hierarchical cluster analysis (HCA) and k-means clustering provided complementary insights into contamination patterns. HCA offered a clear, hierarchical grouping of sites that was valuable for visualising general similarities and identifying broad categories of contamination. In contrast, k-means—drawing on neural network learning principles—yielded higher resolution in cluster formation, allowing subtle differentiation between sites with comparable ΣPBDE levels but distinct compositional signatures. This demonstrated greater sensitivity to variability and enhanced the ability to distinguish among contamination sources and pathways.

By integrating classical statistical techniques with advanced machine-learning approaches, this study demonstrated a methodological framework that strengthens hotspot identification and prioritisation in environmental monitoring. Such combined analyses not only improve the interpretation of current contamination status but also create a foundation for predictive modelling of pollutant dynamics in large, complex river systems. From a management perspective, the identification of hotspots at Neštin, D6, and D10 underscores the need for targeted surveillance, remediation planning, and stricter control of PBDE inputs, particularly from waste and industrial sectors.

Overall, the study shows that bridging traditional multivariate analysis with neural network-based clustering can enhance both the interpretability and the analytical power of sediment contamination assessments. Future work should extend this integrative approach to larger datasets and additional pollutants, enabling more comprehensive risk evaluation and evidence-based decision-making for the protection of riverine ecosystems and human health.

Acknowledgements

This research has been supported by the Ministry of Science, Technological Development and Innovation through Contract No. 451-03-36/2025-03/200156, Fa- culty of Technical Sciences (No.115001-3394/1), HORIZON-MSCA-2021-SE-01, Project No.101086387 (REMARKABLE), and by Ministry of Environmen- tal Protection of the Republic of Serbia and UNDP (proect: Biochar Production from Sewage Sludge: Su- stainable Bioresource Recovery and Carbon Emission Reduction).

REFERENCES

[1] Förstner U., Wittmann G.T.W. Metal Pollution in the Aquatic Environment. Springer, 2012.
[2] Burton G.A. Sediment Toxicity Assessment. CRC Press, 2018.
[3] Eggleton J., Thomas K.V. A review of factors affecting the release and bioavailability of contaminants during sediment disturbance events. Environ. Int., 30 (2004): 973–980.
[4] Darnerud P.O. Toxic effects of brominated flame retardants in man and in wildlife. Environ. Int., 29 (2003): 841–853.
[5] Costa L.G., Giordano G. Developmental neurotoxicity of PBDEs: mechanisms and implications. Environ. Int., 38 (2012): 212–222.
[6] Law R.J., Covaci A. Brominated flame retardants in the environment. Environ. Int., 29 (2003): 725–734.
[7] Brborić M. et al. Unveiling the impact of lipophilic brominated flame retardants in Danube River sediments. EQA — Int. J. Environ. Quality, 71 (2026): 51–60.
[8] Brborić M. et al. Tracking brominated flame retardant patterns in Danube bottom sediment: multivariate statistical approach. Proc. MechEng XXI, 2023.
[9] Jain A.K. Data clustering: 50 years beyond K-means. Pattern Recognition Letters, 31 (2010): 651–666.
[10] Goodfellow I., Bengio Y., Courville A. Deep Learning. MIT Press, 2016.