e-ISSN No.: 2582-4228

Journal of Indian Association for Environmental Management

Journal homepage: www.http://op.niscair.res.in/index/php/JIAEM/index



Assessment of PAHs in Selected Aquatic Systems using GC-MS followed by ANN and PMF Modeling

Soma Mukherjee[1]*, Madhurima Kanji[1], Susmita Bhowmick[1], Soumidipa Chatterjee[1], Anushka Das[1] and Somnath Sen[2&3]

[1]Department of Environmental Science, University of Kalyani, Kalyani, Nadia-741235, West Bengal, India
[2]ICMR-Centre for Ageing and Mental Health, Kolkata-700091, India
[3]ICMR-National Institute of Health Research (NIHR), Bhubaneswar-751023, India
*Corresponding author's Email: somaenvs@klyuniv.ac.in

Submitted:14August 2026Revised:14 October 2026Accepted: 20 October 2026

***Abstract*:Surface water quality is a critical determinant of ecosystem health and public safety, especially in regions experiencing rapid urbanization and industrialization. This study assesses polycyclic aromatic hydrocarbons (PAHs) using triple quadruple gas chromatography-mass spectrometry (GC-MS/MS) in 19 surface water samples collected from selected peri-urban sites in West Bengal. Among 16 polycyclic aromatic hydrocarbons (PAHs), only fluorene (Fl), phenanthrene (Phe), chrysene (Chr) and benzo[a]pyrene (BaP) were detected by GC-MS at some locations. The Water Quality Index (WQI) values were calculated and approximately 10.5% of the samples were classified as poor, 57.9% as very poor and 31.6% under the unsuitable category. Pearson correlation analysis was performed to understand the significant relationships among various parameters. ANN-based prediction and sensitivity analysis revealed that BOD and phenanthrene exerted the greatest influence on the availability of fluorene and chrysene. Positive Matrix Factorization (PMF) indicated three dominant sources such as industrial and vehicular emissions (39%), biomass and biogas burning (35%) and fossil fuel combustion (26%). These results emphasized the urgent need for regular monitoring and effective management to safeguard water resources and public health in rapidly developing regions.**

Keywords: Surface Water, Polycyclic Aromatic Hydrocarbons, Gas Chromatography-Mass Spectrometry, Artificial Neural Network, Positive Matrix Factorization

I. INTRODUCTION

Water plays a vital role in sustaining life and maintaining natural systems. Globally, ensuring surface water quality is essential because it directly affects human healthand ecological balance. However, the contamination load in surface waters have significantly increased due to rapid urbanization, release of poorly treated industrial effluents, and diffuse agricultural runoff that is enriched with nutrients, pesticides, heavy metals, (PAHs), as well as toxic compounds [1-3].

Among key pollutants, polycyclic aromatic hydrocarbons (PAHs) are persistent organic pollutants generated mainly by incomplete combustion of fossil fuels and biomass, as well as petrogenic sources [4]. Low-molecular-weight (2-3 rings) PAHs tend to dominate in the aqueous phase, whereas higher-ring PAHs partition preferentially to sediments and biota. Many PAHs are carcinogenic, mutagenic and endocrine-disrupting in nature [2]. These inputs change physicochemical properties, encourage eutrophication and facilitate bioaccumulation, biomagnification and biotransformation of toxins which ultimately jeopardize the integrity of aquatic ecosystems while increasing the risks to human health [5,6]. Thus, reliable quantification of priority PAHs at environmentally relevant levels is essential for risk assessment and regulatory compliance. High-performance liquid chromatography and gas chromatography coupled to mass spectrometry (GC–MS) are the methods of choice for PAH analysis because of high sensitivity and selectivity, with GC–MS widely preferred for robust structural identification and low (ng/l) level detection in surface waters [3,4].

Also, the identification of pollution sources in peri-urban water bodies requires advanced statistical and modeling approaches due to the coexistence of industrial discharge,

municipal wastewater, traffic emissions, and agricultural runoff. Correlation and multivariate analyses help to explain the relationships between PAHs and physicochemical parameters, revealing the hydrogeochemical controls on their distribution. Positive Matrix Factorization (PMF) enables quantitative apportionment of mixed pollution sources, while Artificial Neural Networks (ANN) provide reliable prediction of PAH concentrations and water quality characteristics from limited physicochemical datasets [7,8]. To date, integrated studies that jointly evaluate GC-MS-resolved PAH fingerprints, conventional physicochemical parameters, multivariate statistics and ANN-PMF-modeled source contributions are limited in peri-urban surface waters [9].

The present study is expected to (i) generate a detailed baseline of dissolved PAH concentrations and profiles in a representative peri-urban aquatic environment; (ii) reveal statistically robust correlations between PAHs and conventional water quality metrics; (iii) identify dominant pollution sources and quantify their relative contributions via ANN, PMF modeling; and (iv) provide targeted pollution control and long term monitoring strategies. Collectively, these outcomes might help to elucidate PAH dynamics in peri-urban watersheds and support more data-driven management of carcinogenic organic pollutants in rapidly changing landscapes.

II. MATERIALS AND METHOD

Water Sampling and Physicochemical Analysis

A total of nineteen surface water samples were randomly collected in precleaned polyethylene bottles (1000 ml) from diverse sites (ponds, ferry ghats, bathing ghat, lakes, canal, car washing center) (Table 1) in February 2024 and analysed for physicochemical parameters (DO, BOD, TH, nitrate, chloride, calcium, magnesium and TA) following APHA methods [10].

The in-situ parameters such as total dissolved solids (TDS), pH and electrical conductivity (EC) were measured immediately at the site with the EUTECH multi-parameter tester 35 series. Dissolved oxygen (DO) was determined by the Winkler iodometric titration method, while biochemical oxygen demand (BOD_5) was analysed using the 5-day incubation method at 20 °C. The total hardness (TH), calcium (Ca^{2+}) and magnesium (Mg^{2+}) were estimated using EDTA (ethylene diamine tetraacetic acid) complexometric titration method. The total alkalinity (TA) was measured by acid titration (using HCl, methyl orange and phenolphthalein). Argentometric titration was used to estimate the chloride (Cl^-) content (using $AgNO_3$ and K_2CrO_4) present in surface water samples. UV-Vis spectrophotometer (SHIMADZU-1900) was used to determine the nitrate (NO_3^-) concentration present in water samples [11]. The map (ArcGIS 10.6, Google Earth Pro) for sampling locations was shown in the Figure 1.

Water Quality Index (WQI)

Water quality was evaluated using the Weight Arithmetic Water Quality Index (WAWQI), which classifies water into five quality categories (Excellent: 0-25, Good: 26-50, Poor: 51-75, Very Poor: 76-100, Unsuitable for drinking purposes: >100) [12,13].

To evaluate WQI, the unit weight was given by:

$$W_i = \frac{K}{S_i} \quad \text{.....(1)}$$

where S_i= standard permissible limits for each parameter and

$$K = \left(\sum \frac{1}{S_n}\right)^{-1} \quad \text{..... (2)}$$

The quality Q_i rating was given by:

$$Q_i = \left(\frac{C_i - V_{ideal}}{S_i - V_{ideal}}\right) \times 100 \quad \text{..... (3)}$$

Where, C_i = concentration of the parameter and S_i = standard permissible limit of the parameter. The following ideal values (V_{ideal}) were employed: 7.0 for pH, 14.6 mg/l (saturation level at 0 °C) for dissolved oxygen (DO) and 0 for all other parameters. The sub-index was given by:

$$W_i \times Q_i \quad \text{..... (4)}$$

and the Water Quality Index (WQI) was given by:

$$WQI = \frac{\sum W_i Q_i}{\sum W_i} \quad \text{..... (5)}$$

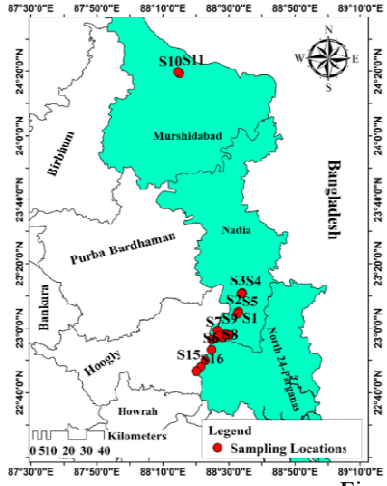


Table 1
Sampling locations with geographical coordinates.

Sampling Sites No.	Latitude	Longitude	Description
S1	23.07663°N	88.54441°E	Pond near Tyre dealer centre, Chhatimtala; Chakdaha
S2	23.08389°N	88.54776°E	Pond near Car washing centre, Chakdaha
S3	23.18272°N	88.56616°E	Ghat of Churni River, Ranaghat for idol immersion
S4	23.17949°N	88.56186°E	Bathing ghat of Churni River, Ranaghat
S5	23.07447°N	88.54281°E	Pond near Highway NH-12, Chakdaha
S6	22.95545°N	88.47177°E	Bikrampur ghat (Kulia jhill)
S7	22.96949°N	88.47937°E	Pond near TDK industry
S8	22.9725°N	88.4433°E	Kalyani lake
S9	22.98736°N	88.4418°E	Pond near Sati mayerbari
S10	24.32124°N	88.25045°E	Sundarpur (canal near river)
S11	24.33053°N	88.24158°E	Sundarpur (pond)
S12	22.94044°N	88.41375°E	Halisahar ferry ghat
S13	22.88938°N	88.41175°E	Naihati ferry ghat
S14	22.83118°N	88.37947°E	Shyamnagar ghat
S15	22.80002°N	88.35742°E	Shyamnagar ferry ghat
S16	22.77757°N	88.33333°E	North Barrackpore ferry ghat
S17	22.95168°N	88.46244°E	Kulia jhill
S18	22.96479°N	88.50178°E	Dogachhina (road side lake)
S19	22.96166°N	88.44813°E	Kalyani Block B (car washing)

Figure 1: Map indicating sampling sites

Sample Extraction and GC-MS/MS Analysis

By grab sampling, 3 L of surface water was collected in pre-treated amber glass bottles fitted with a screw cap lined with PTFE. Water samples were filtered to remove suspended particulates and transferred into a separatory funnel for liquid–liquid extraction using 15 ml of n-hexane. The samples were mixed on an orbital shaker for 30 minutes and the aqueous layer was drained, with the hexane layer collected and the extraction was repeated twice more to improve recovery. The combined extracts were cleaned and reconstitute by solid-phase extraction (SPE) using a C18 cartridge (preconditioned with 10 ml each of dichloromethane, methanol and Milli-Q water). Analytes were eluted with 10 ml dichloromethane, concentrated to ~0.5 ml using a rotary evaporator (35°C) and nitrogen stream, spiked with 20 µl of internal standards and adjusted with dichloromethane in 1.0 ml vial for Gas chromatography-tandem mass spectrometry (GC–MS/MS) analysis, where 16 USEPA priority PAHs were quantified using a external certified reference standard mixture (Brand: Supalco, catalog no.49156 EPA TCL PAH Mix certified reference material, in acetonitrile: methanol (9:1) , lot No LRAD7500 [14,15]. GC-MS/MS) analysis was performed on a Shimadzu GCMS-TQ8040 system with an AOC-20i+s autosampler and an SH-Rtx-5Sil MS column (30 m × 0.25 mm ID × 0.25 µm film thickness). A 1 µl sample was injected in splitless mode at 300 °C with high-pressure injection (250 kPa, 2 min). Helium was the carrier gas in constant linear velocity mode (50.0 cm/s; purge flow 3.0 ml/min). The oven was held at 110 °C (5 min), ramped at 25 °C/min to 250 °C (5 min), then at 5 °C/min to 290 °C (5 min) and 320 °C (5 min), for a 39.60 min run. Ion source and interface temperatures were 230 °C and 300 °C, with a 3 min solvent cut. Data were acquired in multiple reaction monitoring (MRM) mode. Each of the 16 PAH was monitored via one quantitative MRM transition and two qualitative MRM transitions. The details of MRM transitions and collision energy (CE) values grouping are tabulated in Supplementary Table S-3.

Diagnostic Ratio

A ΣLMW/ΣHMW ratio was employed for source identification (Table S-4). The ratio<1 indicated a larger contribution of HMW PAHs and may suggest pyrogenic sources, ~1 indicated combustion, and a ratio >1 indicated a predominance of LMW PAHs and is typically linked with petrogenic sources [16,17].

Statistical Analysis

Pearson's correlation coefficient (r) was also used (RStudio version 2024.04.2.0) to evaluate the linear relationships among physicochemical parameters and four PAHs found in water samples. Pearson's correlation coefficient ranged from −1 to 1, and different values quantified the degree of interdependence between parameters [18].

$$r = \frac{n\sum xy \ \ (\sum x)(\sum y)}{\sqrt{[n\sum x^2-(\sum x)^2][n\sum y^2-(\sum y)^2]}} \quad \text{.....(6)}$$

Where, r is the Pearson correlation coefficient and two water quality variables are x and y.

ANN Modeling

A multilayer feed-forward perceptron (MLP) ANN Modeling (SPSS 20.0) was applied to understand nonlinear processes as given in equation 7 [19 and references therein]:

$$y_k = f_0\left[\sum_{i=1}^{M_N} W_{kj} \times f_h\left(\sum_{i=1}^{N_N} W_{ji} \times X_i + \omega_{j0}\right) + \omega_{k0}\right] \text{.....(7)}$$

Here, W_{ji} represents the weight connecting the i^{th} input neuron to the j^{th} hidden neuron, with ω_{j0} as its bias and f_h as the activation function. Similarly, W_{kj} links the j^{th} hidden neuron to the k^{th} output neuron, with ω_{k0} as the output bias and f_0 as its activation function. X_i denotes input parameters, y_j the calculated output, while M_N and N_N indicate the number of input and hidden neurons, respectively.

The Normalized value (X_N) before training and testing was given by equation 8

$$X_N = \frac{X - X_{min}}{X_{max} - X_{min}} \text{.....(8)}$$

Here, X denotes the original value, $X_{max}$and X_{min} are the maximum and minimum values of the dataset.The ANN performance was assessed using following equations:

a. The coefficient of determination:

$$R^2 = \frac{[\sum_{i=1}^{N}(P_i-\bar{P})(O_i-\bar{O})]^2}{[\sum_{i=1}^{N}(P_i-\bar{P})^2(O_i-\bar{O})^2]} \quad \text{.....(9)}$$

b. Root mean squared error (RMSE):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(P_i-O_i)^2}{N}} \quad \text{..... (10)}$$

c. Mean squared error (MSE):

$$MSE = \frac{\sum_{i=1}^{N}(P_i-O_i)^2}{N} \quad \text{..... (11)}$$

Here, N is the number of observations; P_i represents the predicted (forecast) values, O_i the observed values, $\bar{P}$ and $\bar{O}$ are their respective means.

Positive Matrix Factorization (PMF) Modeling

The PMF (USEPA PMF 5.0) model was applied to identify the dominant sources of PAHs in surface water samples [20-24] as follows:

$$X = GF + E \text{.....(12)}$$

In the PMF model, the water quality dataset X (n×m; n = samples, m = chemical species) is decomposed into two matrices, G (n×p; p = sources) and F (p×m), with a residual matrix E (n×m) as given in Eq.12.

$$x_{ij} = \sum_{k=1}^{p} g_{ik} f_{kj} + e_{ij} \quad \text{.....(13)}$$

Here, x_{ij} denotes the concentration of the j^{th} parameter in the i^{th} sample; g_{ik} is the contribution of the k^{th} source to the i^{th} sample; and f_{kj} represents the concentration of the j^{th} parameter in the k^{th}source (Eq. 13). The model applies non-negativity constraints ($gik \geq 0, fkj \geq 0$).

The residual matrix e_{ij} was derived by minimizing the objective function:

$$Q = \sum_{i=1}^{n}\sum_{j=1}^{m}\left(\frac{e_{ij}}{u_{ij}}\right)^2 \quad \text{.....(14)}$$

Here, u_{ij} denotes the uncertainty of parameter i in the j^{th} sample (Eq. 14). Uncertainty was estimated using the method detection limit (MDL) and the standard deviation (SD) of surrogate standards, with a 5% relative SD applied in this study.

For concentrations ≤ MDL, uncertainty was calculated by Eqs 15 and 16:

$$\mu = \frac{5}{6} \times MDL \quad \text{.....(15)}$$

otherwise:

$$\mu = \sqrt{(\sigma + c)^2 + (0.5 \times MDL)^2} \quad \text{.....(16)}$$

where μ denotes uncertainty, σ the relative SD, and c the parameter concentration.

III. Results and Discussion

Physicochemical Data

Surface water samples collected from 19 sites were analysed for different physicochemical parameters and compared with WHO and CPCB standards (Table S-1) [25,26].The pH ranged from 6.3 to 8.3, with few samples (S3 and S6) marginally below the acceptable limit of 6.5–8.5. Electrical conductivity (432–845µS/cm) and TDS (332.8–613.6 mg/l) were found to lie within acceptable limits (EC: 500-1500µS/cm and TDS: 500-1000 mg/l, WHO, 2017). The calcium (28.82–145.85 mg/l) and magnesium (8.04–115.81 mg/l) ion concentrations remained within the specified limits of WHO (75-200 mg/l) and chloride levels (36.92–237.1 mg/l) were also below the permissible threshold limit (>250 mg/l). DO values varied from 5.3 mg/l to 9.0 mg/l (CPCB limit: 5.0 mg/l), whereas BOD ranged from 6.0 to 13.0 mg/l (CPCB limit: 3.0 mg/l), with most samples exceeding the acceptable BOD level. Such deviations may be attributed to inputs from untreated domestic wastewater and other anthropogenic influences. Major cation and anion distributions were illustrated in the Figure 2.

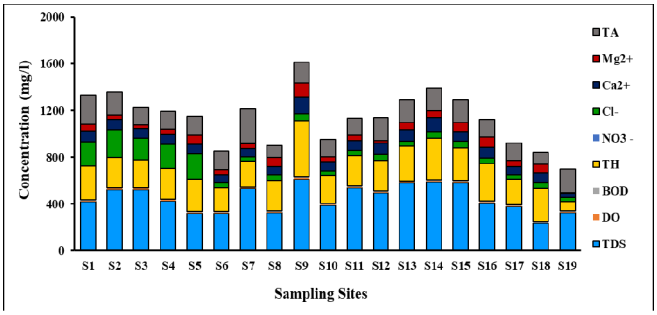


Figure 2: Physicochemical parameters of surface water samples at various sampling sites

Water quality index (WQI)

The water quality index represents the overall quality of groundwater whether suitable or not for drinking purposes. Nineteen samples from various sites ranged from "poor" (51-75) to "unsuitable," categories with WQI>100 for five samples. Approximately 10.5% of the samples were classified as poor, 57.9% as very poor and 31.6% as unsuitable (Table 2). Elevated WQI scores were possibly due to high BOD values.

PAHs Analysis

The GC-MS was used for PAHs separation, identification and quantification of USEPA priority sixteen PAHs, *viz.* Naphthalene, Acenaphthene, Acenaphthylene, Fluorene, Phenanthrene, Anthracene, Fluoranthene, Pyrene, Benzo(a)anthracene, Chrysene, Benzo(b)fluoranthene, Benzo(k)fluoranthene, Benzo(a)pyrene, Indeno(1,2,3-cd)pyrene, Dibenz(a,h)anthracene and Benzo(g,h,i) perylene. The targeted PAHs compounds, number of rings, molecular weight & formula, retention time, LOD &LOQ in the instrument were summarised in the Table S-2.

It was found that fluorene (Fl), phenanthrene (Phe), chrysene (Chr) and benzo[a]pyrene (BaP) compounds were detected in several samples. Their concentrations ranged from non-detectable (ND) to detectable levels as follows: Fl (ND–10.6 ng/l), Phe (ND–2.95 ng/l), Chr (ND–1.44 ng/l) and BaP (ND–3.25 ng/l), all of which remained within permissible limits (Table 2). The concentrations of these four PAHs at the different sampling locations were shown in the Figure 3a. The frequency of occurrence was highest for Fl (57.89%) followed by Chr (36.84%), Phe (21.05%) and BaP (10.53%) (Figure 3b). The concentrations of PAHs in the present study were compared with reported riverine and estuary systems throughout India (Table S-4) and difference was observed due to different hydrological conditions and other anthropogenic activity.

The distribution and behavior of PAHs in the studied region were strongly influenced by the prevailing hydrochemical conditions. Moderate DO and BOD, combined with low total hardness, enhanced the solubility and mobility of PAHs. Under these conditions, low-molecular-weight PAHs such as Fl became more available due to their higher solubility in soft water and was degraded more rapidly through aerobic processes. In contrast, high-molecular-weight PAHs (Chr, BaP) persisted because of their strong hydrophobicity and slower degradation rates. The presence of these PAHs in the studied region likely resulted from non-point pollution sources, including urban runoff, industrial effluents, vehicular emissions, biomass burning and fossil-fuel combustion. The simultaneous detection of Fl, Phe, Chr and BaP also indicated possible co-solubilization promoted by natural surfactants, humic matter and organic matter, which facilitated the transport of hydrophobic PAHs.

TABLE 2
Concentrations of PAHs by GC-MS and WQI at different sampling locations

Sampling Sites	Nap	Acy	Ace	Fl	Phe	Ant	Flu	Pyr	BaA	Chr	BbF	BkF	BaP	DBahA	BghiP	InP	WQI
	Conc. (ng/l)																
S1	ND	ND	ND	10.6	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	89.0
S2	ND	ND	ND	2.7	ND	ND	ND	ND	ND	0.18	ND	ND	ND	ND	ND	ND	86.7
S3	ND	ND	ND	1.6	ND	ND	ND	ND	ND	ND	ND	ND	0.02	ND	ND	ND	93.2
S4	ND	ND	ND	1.8	ND	ND	ND	ND	ND	0.18	ND	ND	ND	ND	ND	ND	77.5
S5	ND	ND	ND	2.5	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	77.0
S6	ND	ND	ND	0.04	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	93.4
S7	ND	ND	ND	2.7	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	106.5
S8	ND	ND	ND	6.1	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	112.3
S9	ND	ND	ND	3.5	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	130.9
S10	ND	ND	ND	0.4	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	84.0
S11	ND	ND	ND	ND	0.33	ND	ND	ND	ND	0.34	ND	ND	ND	ND	ND	ND	94.8
S12	ND	ND	ND	ND	0.88	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	78.4
S13	ND	ND	ND	0.10	2.95	ND	ND	ND	ND	1.44	ND	ND	ND	ND	ND	ND	69.8
S14	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	3.25	ND	ND	ND	67.8
S15	ND	ND	ND	ND	ND	ND	ND	ND	ND	0.67	ND	ND	ND	ND	ND	ND	93.6
S16	ND	ND	ND	ND	0.11	ND	ND	ND	ND	0.82	ND	ND	ND	ND	ND	ND	104.3
S17	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	98.9
S18	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	109.1
S19	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	131.5
USEPA[27]	-	-	-	-	-	-	-	-	100	200	200	200	200	300	-	400	-

[Note: Naphthalene(Nap), Acenaphthene (Ace), Acenaphthylene (Acy), Fluorene (Fl), Phenanthrene (Phe), Anthracene (Ant), Fluoranthene (Flu), Pyrene (Pyr), Benzo[*a*]anthracene (BaA), Chrysene (Chr), Benzo[*b*]fluoranthene (BbF), Benzo[*k*]fluoranthene (BkF), Benzo[*a*]pyrene (BaP), Dibenz[*a*,*h*]anthracene (DahA), Benzo[g, h,i] perylene (BghiP), and Indeno(1,2,3-*cd*)pyrene (IND)]

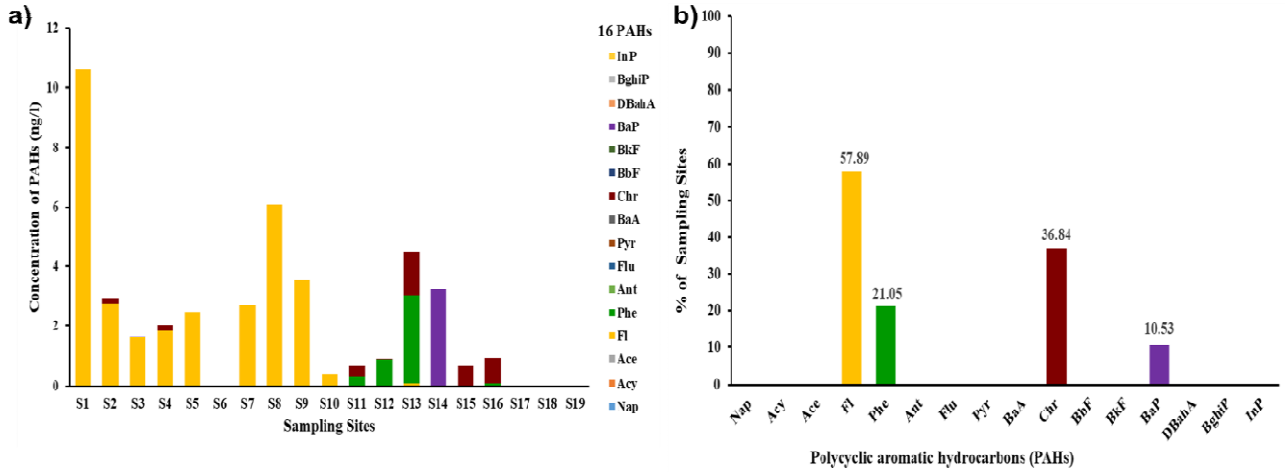


Figure 3: a) Concentration of 16 PAHs (USEPA priority PAHs) at different sampling sites of the studied region, b) PAHs frequency with percentage of sampling sites

Diagnostic Ratio

The ratio of ΣLMW/ΣHMW was employed to evaluate potential sources of PAH. The ratio could only be calculated for S11 (0.97) and S13 (2.05) (where both LMW and HMW PAHs were quantifiable). Maximum concentrations of LMW and HMW PAHs in the samples were either below detection limits (<LOD) or absent (ND) (Table S-5). Whereas the ratio in Sundarpur (pond) (S11) was almost equal to 1, indicating combustion, the prevalence of LMW PAHs in sampling site Naihati ferry ghat (S13) indicated petrogenic impact.

Statistical Analysis

The Pearson correlation analysis revealed a strong positive relationship between Ca^{2+} and Mg^{2+} (r = 0.58) and EC with TDS (r=0.93), suggesting a shared geogenic origin. Conversely, a moderate negative correlation was observed between DO and BOD (r = -0.50), indicating oxygen depletion in water with higher organic loads [28].

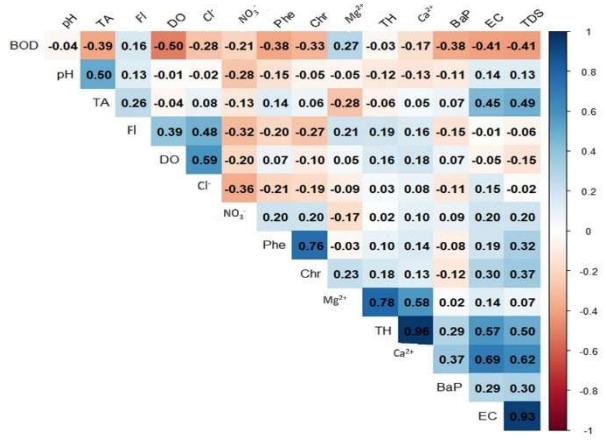


Figure 4: Correlations between physicochemical parameters and four PAHs present in different surface water samples

The correlation diagram (Figure 4) revealed interactions between physicochemical parameters and PAHs (fluorene, phenanthrene, chrysene, and benzo[a]pyrene). Fl exhibited significant as well as moderate positive correlations with DO (r = 0.39) and Cl^- (r = 0.48), indicating linkage with oxygenated and mineral-rich conditions. Phe showed a strong correlation with Chr (r = 0.76), suggesting a common source such as combustion or petroleum inputs. Chr was positively related to EC (r = 0.30) and TDS (r = 0.37), reflecting its association with ionic pollution. BaP showed weak to moderate correlations with TDS (r = 0.30), Ca^{2+} (r = 0.37), EC (r=0.29) and TH (r = 0.29), implying anthropogenic influence. Overall, the intercorrelations indicated that PAHs might have shared common sources and were influenced by the hydrochemical characteristics of the water [29-32].

ANN Modeling

Physicochemical data along with PAHs from 19 sites were used for ANN modeling (70% training, 30% testing). The optimal network was identified by evaluating various transfer functions and layer configurations to select the model with the highest R^2 and the lowest errors (relative error, MSE, RMSE). Biases and weights were updated iteratively and model generalization was validated using separate datasets. A back propagation algorithm was applied to relate the input variables to the output (Fl and Chr were considered due to their availability at most sampling locations). The network specifications were provided in Table S-6 and Table S-7 (Figures S-1 and S-2).

TABLE 3
The statistical results of the ANNs model

		MSE	RMSE	R-Squar
Fl	Training	0.0003	0.018	0.99
	Testing	0.024	0.154	0.93
Chr	Training	0.038	0.194	0.92
	Testing	0.017	0.131	0.92

Model performance (Table 3) showed low MSE & RMSE values and high R^2 values for both Fl and Chr in the training (Fl: 0.0003, 0.018, 0.99; Chr: 0.038, 0.194, 0.92) and testing phases (Fl: 0.024, 0.154, 0.93; Chr: 0.017, 0.131, 0.92). These results confirmed the accuracy of the ANN model, with eight hidden neurons producing minimal errors. The representative

comparison plot of Fl (Figure 5a and Figure 5b) demonstrated strong agreement between measured and predicted values, indicating that the ANN models reliably estimated Fl (Figure 5c and Figure 5d for Chr) concentrations from simple input parameters.

Figure 6a shows that BOD had the greatest influence on Fl prediction in the ANN model sensitivity analysis, followed by DO and TA, suggesting low persistence of Fl in the aqueous environment. Phe showed a significant influence on the availability of Chr due to their similar pollution sources, transport behaviour and environmental fate in water, followed by the other input variables (Figure 6b).

PMF Modeling

PMF modeling identified three main factors (Figure 7), each representing a distinct source of the four PAHs in the study area. In random-start mode, PMF was run with 3–10 factors and evaluated using Q(robust) vs. Q(true), with 100 base runs. Three factors were selected based on Q(robust) ≈ Q(true) and strong prediction–observation correlations [21]. The Q(robust)/Q(true) ratio of 0.9 and the average R^2 value of 0.77 for the four indicators confirmed the suitability of this factor number. The residuals of almost all samples varied between -3 and 3.

In the study area, the surface water showed three factors: F1 (26%), dominated by Phe, Chr, and Fl; F2 (39%), the major contributor, dominated by Chr, BaP and Phe; and F3 (35%), dominated by Flu, BaP and Phe. In F1, Phe dominated (82.7%), with lower contributions from Chr (19.5%) and Flu (1.9%). Phe largely derived from fossil fuel combustion, while BaP and Flu were mainly associated with pyrogenic inputs such as coal and biomass burning. Hence, F1 was attributed to fossil fuel combustion. In F2, Chr (80.5%) and BaP (64.6%) were the dominant contributors, followed by Phe (11.7%), indicating that the major inputs originated from high-temperature combustion (coal, biomass, vehicular emissions), petroleum discharges, and industrial effluents. In F3, Fl (98.1%) and BaP (35.4%) dominated, followed by Phe (5.6%), suggesting major inputs from petroleum sources and the combustion of biomass, coal, natural gas, biogas and vehicular emissions.

TABLE 4
Factor profiles obtained from PMF model

Parameter	Conc. of Species			% of Species		
	F1	F2	F 3	F1	F2	F 3
Fluorene (Fl)	0.032	0	1.624	1.9	0	98.1
Phenanthrene (Phe)	0.226	0.032	0.015	82.7	11.7	5.6
Chrysene (Chr)	0.049	0.202	0	19.5	80.5	0
Benzo[a]pyrene (BaP)	0	0.006	0.003	0	64.6	35.4

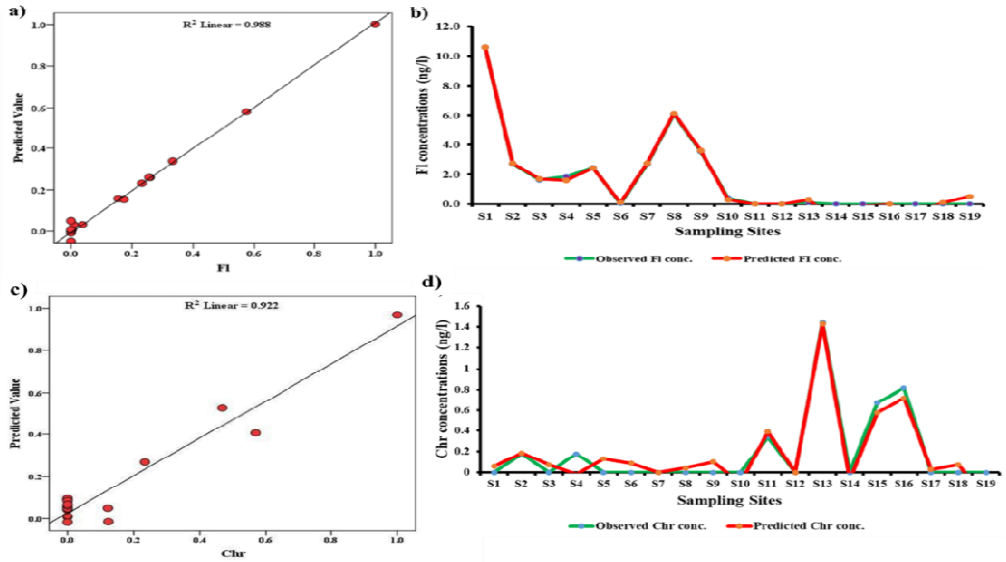


Figure 5: The comparison plots: a) Scatter plot of computed and predicted Fl, b) Line diagram of observed and predicted Fl concentration, c) Scatter plot of computed and predicted Chr, d) Line diagram of observed and predicted Chr concentration

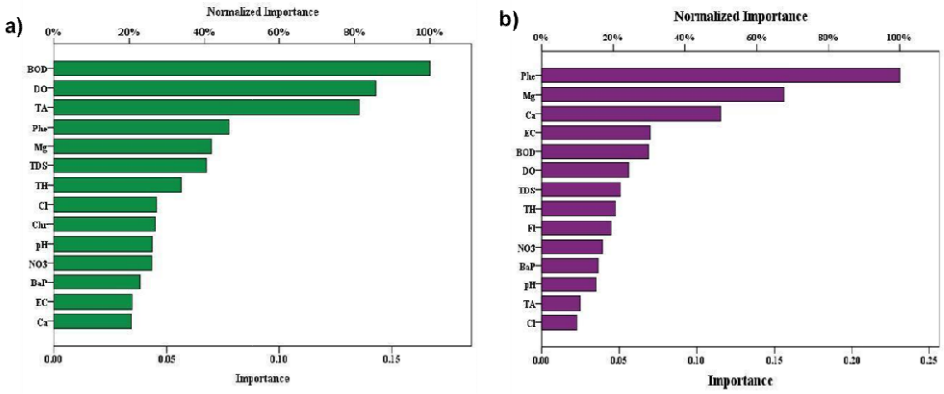


Figure 6: Influence of input variables on a) Fl and b) Chr

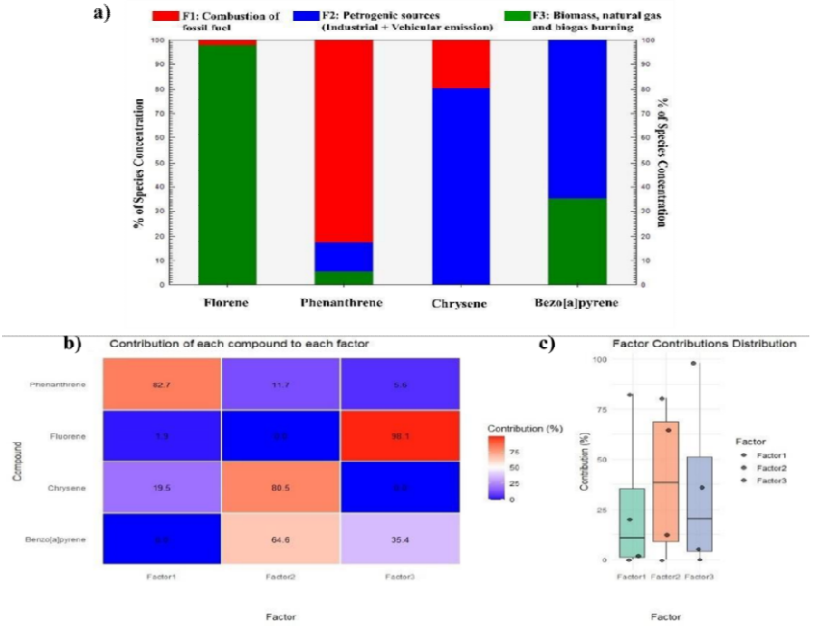


Figure 7: a) Contribution factor fingerprints from PMF model, b) Contribution of each compound to each factor, c) Average contribution of pollution sources

Based on the source contribution rates from the PMF model (Table 4), the contributions of the three factors were converted to relative rates and the contribution of each factor was presented in the Fig. 7. Thus, the PMF model identified three main factors (F1, F2, F3) contributing to surface water contamination in the study area, and field observations around the sampling sites were used to understand and validate the identified sources (Table S-8). The main contributions were F2 and F3, which are associated with combustion, petroleum, and industrial sources, whereas F1 had a relatively less impact. These dominant factors highlighted the strong influence of anthropogenic activities, particularly combustion and petroleum-related sources on the availability of PAHs.

IV. Conclusion

The present study provides a comprehensive assessment of surface water quality in selected peri-urban locations by using physicochemical parameters and the Water Quality Index (WQI). The results showed that the water quality status varied from poor to unacceptable across selected sites due to anthropogenic influence, with values ranging from 67.8 to 131.5. Fluorene, phenanthrene, chrysene and benzo[a]pyrene were the only four of the 16 priority PAHs detected in selected samples. The BOD and phenanthrene were found to be predominant factors impacting the distribution of fluorene and chrysene as obtained from the sensitivity analysis in ANN modeling. The PMF identified three components that explained the difference in percentage of variance, primarily attributing PAHs to vehicle emissions and industrial discharge (39%), natural and biogas burning (35%) and fossil fuel combustion (26%). Overall, the combination of WQI evaluation, statistical analysis, ANN-based predictive modeling and PMF source apportionment within an integrated framework strengthened the understanding of hydrochemical processes and pollution sources in peri-urban aquatic systems. Data-driven and receptor-based combination strategies provide a dependable foundation for site-specific management, pollution mitigation and sustainable water resource planning in rapidly urbanizing transitional regions.

V. REFERENCES

Y. Liu, Y. He, Y. Liu, H.J. Liu, S. Tao, W.X. Liu, Source identification and ecological risks of parent and substituted polycyclic aromatic hydrocarbons in river surface sediment–pore water systems: effects of multiple factors, Sci. Total Environ. 858 (2023) 159921, https://doi.org/10.1016/J.SCITOTENV.2022.159921.

H. Haghnazar, Y. Abbasi, R. Morovati, K.H. Johannesson, R. Somma, M. Pourakbar, E. Aghayani, Polycyclic aromatic hydrocarbons (PAHs) in the surficial sediments of the Abadan freshwater resources – northwest of the Persian Gulf, J. Geochem. Explor. 258 (2024) 107390, https://doi.org/10.1016/J.GEXPLO.2024.107390.

R.M.M. Muhsin, T.S.B. Abd Manan, J. Bidai, M.S.A. Mangat, Z.M. Hanafiah, A. Gohari, N. Ahmad, F. Ahmad, S. Beddu, N.L.M. Kamal, D. Mohamad, S.A.H. Aldala'in, M.R. Ul Mustafa, W.H.M. Wan Mohtar, M.H. Isa, M.S. Yusoff, H. Abdul Aziz, Polycyclic aromatic hydrocarbons (PAHs) occurrences in water bodies, extraction techniques, detection methods, and standardized guidelines for PAHs in aqueous solutions, Sci. Total Environ. 972 (2025) 179123, https://doi.org/10.1016/j.scitotenv.2025.179123.

R.A. Grmasha, M.H. Abdulameer, C. Stenger-Kovács, O.J. Al-Sareji, Z. Al-Gazali, R.A. Al-Juboori, M. Meiczinger, K.S. Hashim, Polycyclic aromatic hydrocarbons in the surface water and sediment along Euphrates River system: occurrence, sources, ecological and health risk assessment, Mar. Pollut. Bull. 187 (2023) 114568, https://doi.org/10.1016/j.marpolbul.2022.114568.

S. Barua, B. P. Mukhopadhyay, A. Bera, Hydrochemical assessment of groundwater for irrigation suitability in the alluvial aquifers of Dakshin Dinajpur district, West Bengal, India, Environmental Earth Sciences. 80(16) (2021) 1–14, 10.1007/s12665-021-09832-y.

D. Yu, J. Deng, Q. Jiang, H. Liu, C. Yu, H. Ma, S. Pu, Evaluation of groundwater quality with multi-source pollution based on source identification and health risks, Science of the Total Environment. 949 (2024) 175064, 10.1016/j.scitotenv.2024.175064.

S. C. Pal, D. Ruidas, A. Saha, A. R. M. T. Islam, I. Chowdhuri, Application of novel data-mining technique-based nitrate concentration susceptibility prediction approach for coastal aquifers in India, Journal of Cleaner Production. 346 (2022) 131205, 10.1016/j.jclepro.2022.131205.

Y. C. Cho, H. Choi, S. J. Yu, S. H. Kim, J. K. Im, Assessment of spatiotemporal variations in the water quality of the Han River Basin, South Korea, using multivariate statistical and APCS-MLR modeling techniques, Agronomy. 11 (2021) 2469, 10.3390/agronomy11122469.

T. Hasrod, Y.B. Nuapia, H. Tutu, Comparison of individual and ensemble machine learning models for prediction of sulphate levels in untreated and treated Acid Mine Drainage, Environ. Monit. Assess. 196 (2024) 332, https://doi.org/10.1007/s10661-024-12467-8.

R.B. Baird, A.D. Eaton, E.W. Rice (Eds.), Standard Methods for the Examination of Water and Wastewater, 23rd ed., American Public Health Association, American Water Works Association, Water Environment Federation, Washington, DC, USA, (2017).

R.K. Trivedy, P.K. Goel, Chemical and biological methods for water pollution studies, Environ. Publ. (Karad, India) 6 (1986) 10–12.

R.M. Brown, N.J. McCleiland, R.A. Deiniger, M.F. O'Connor, Water quality index crossing the physical barrier, in: S.H. Jenkins (Ed.), Proceedings of the International Conference on Water Pollution Research, Jerusalem, 6 (1972) 787–797.

M. Hagage, A. G. A. Hewaidy, A. M. Abdulaziz, Groundwater quality assessment for drinking, irrigation, aquaculture, and industrial uses in the waterlogged northeastern Nile Delta, Egypt, Modeling Earth Systems and Environment. 11(1) (2025), 10.1007/s40808-024-02242-6.

APHA, AWWA, WEF, 6410 B. Polycyclic aromatic hydrocarbons (PAHs) by GC/MS, in: Standard Methods for the Examination of Water and Wastewater, 23rd ed., American Public Health Association, Washington, DC, (2017).

APHA, AWWA, WEF, 6440 B. Polycyclic aromatic hydrocarbons (PAHs) by HPLC, in: Standard Methods for the Examination of Water and Wastewater, 23rd ed., American Public Health Association, Washington, DC, (2017).

M. Tobiszewski, J. Namieśnik, PAH diagnostic ratios for the identification of pollution emission sources, Environ. Pollut. 162 (2011) 110–119, https://doi.org/10.1016/j.envpol.2011.10.025

W. Zhang, S. Zhang, C. Wan, D. Yue, Y. Ye, X. Wang, Source diagnostics of polycyclic aromatic hydrocarbons in urban road runoff, dust, rain and canopy throughfall, Environ. Pollut. 153 (2008) 594–601, https://doi.org/10.1016/j.envpol.2007.09.004

P. Y. Kharat, M. M. Nair, P. S. Rakesh, & C. K. Haridevi, Distribution and bioaccumulation status of polycyclic aromatic hydrocarbons (PAHs) in Veraval coastal waters using copepods as bio-indicators, Environmental Monitoring and Assessment, 196(8) (2024) 711, https://doi.org/10.1007/s10661-024-12805-w

L.K. Al-Waeli, J.H. Sahib, H.A. Abbas, ANN-based model to predict groundwater salinity: a case study of West Najaf–Kerbala region, Open Eng. 12 (1) (2022) 120–128, https://doi.org/10.1515/eng-2022-0025.

P. Paatero, U. Tapper, Positive matrix factorization: a non-negative factor model with optimal utilization of error estimates of data values, Environmetrics 5 (2) (1994) 111–126.

G.A. Norris, R. Duvall, S. Brown, S. Bai, EPA Positive Matrix Factorization (PMF) 5.0 Fundamentals and User Guide, U.S. EPA, 2014.

Q. Zhang, J. Zhang, H. Wang, T. Zhai, L. Liu, G. Li, Z. Xu, Spatial patterns in water quality and source apportionment in a typical cascade development river southwestern China using PMF modeling and multivariate statistical techniques, Chemosphere 311 (2022) 137139, https://doi.org/10.1016/j.chemosphere.2022.137139.

H. Shi, Y. Du, Y. Li, Y. Deng, Y. Tao, T. Ma, Determination of high-risk factors and related spatially influencing variables of heavy metals in groundwater, J. Environ. Manag. 358 (2024) 120853, https://doi.org/10.1016/j.jenvman.2024.120853.

D. Yu, J. Deng, Q. Jiang, H. Liu, C. Yu, H. Ma, S. Pu, Evaluation of groundwater quality with multi-source pollution based on source identification and health risks, Sci. Total Environ. 949 (2024) 175064, https://doi.org/10.1016/j.scitotenv.2024.175064.

World Health Organization, Guidelines for Drinking-water Quality, fourth ed., incorporating the first addendum, WHO, Geneva, Switzerland, (2017).

Central Pollution Control Board (CPCB), Water Quality Criteria and Classification of Inland Surface Waters, Ministry of Environment, Forest and Climate Change, Government of India, New Delhi, (2019).

United States Environmental Protection Agency, Ambient Water Quality Criteria for Polynuclear Aromatic Hydrocarbons (EPA 440/5–80-069), US EPA, Washington DC, (1980).

X. Qin, H. Wang, J. Gong, Y. Ye, K. Zhou, N. Xu, L. Li, J. Li, Multivariate statistics and hydrochemistry combined to reveal the factors affecting shallow groundwater evolution in a typical area of the Huaibei Plain, China, Water 17 (2025) 962, https://doi.org/10.3390/w17070962.

K. Zhang, S. Chang, Q. Zhang, Y. Bai, E. Wang, Y. Fan, X. Tu, Q. Fu, L. Wei, Y. Yu, Occurrence, source modeling, influencing factors and exposure assessment of polycyclic aromatic hydrocarbons in water sources: A mega-study from mainland China, Environ. Technol. Innov. 35 (2024) 103634, https://doi.org/10.1016/j.eti.2024.103634.

P.Y. Kharat, M.M. Nair, P.S. Rakesh, C.K. Haridevi, Distribution and bioaccumulation status of polycyclic aromatic hydrocarbons (PAHs) in Veraval coastal waters using copepods as bio-indicators, Environ. Monit. Assess. 196 (2024) 711, https://doi.org/10.1007/s10661-024-12805-w.

A. Baran, M. Tarnawski, K. Urbański, A. Klimkowicz-Pawlas, I. Spałek, Concentration, sources and risk assessment of PAHs in bottom sediments, Environ. Sci. Pollut. Res. 24 (2017) 23180–23195, https://doi.org/10.1007/s11356-017-9944-y.

K. Khosravi, A.A. Farooque, A.R. Shahvaran, P. Daggupati, S. Heddam, J. Hatamiafkoueieh, Beyond conventional modeling: A cutting-edge hybrid IAER-AMT decision-tree-based algorithm for high-resolution river turbidity prediction, Ain Shams Eng. J. 16 (2025) 103511, https://doi.org/10.1016/j.asej.2025.103511.

VI. Supplementary Information

TABLE S-1
Physicochemical parameters of surface water samples

Sampling Sites No.	pH	EC (µS/cm)	TDS (mg/l)	DO (mg/l)	BOD (mg/l)	TH (mg/l)	NO_3^- (mg/l)	Cl^- (mg/l)	Ca^{2+} (mg/l)	Mg^{2+} (mg/l)	TA (mg/l)
S1	6.89	635	418.4	8.1	9	288	0.07	207.3	92.6	56.78	250
S2	8.3	801	520.2	9	6.4	260	0.07	237.1	88.8	38.27	200
S3	6.3	710	522.1	6.59	10	236	0.02	188.8	82	31.25	150
S4	6.7	639	424.6	8.5	8	260	0.05	207.3	85.6	46.26	150
S5	6.7	538	316.8	8.7	8	280	0.05	218.68	80.02	80.19	160
S6	6.3	536	317.4	7.2	10.4	200	0.03	51.12	62.04	45.09	160
S7	8.3	733	528.9	7	8.2	220	0.09	38.17	71.43	41.64	300
S8	7.1	432	320.8	8.3	12	260	0.05	45.09	72.25	79.59	100
S9	7.3	845	613.6	6.2	13	480	0.04	56.8	145.85	115.81	180
S10	6.8	540	387.3	7.2	8.1	240.6	0.96	37	80.8	38.84	154.2

Sampling Sites No.	pH	EC (µS/cm)	TDS (mg/l)	DO (mg/l)	BOD (mg/l)	TH (mg/l)	NO_3^- (mg/l)	Cl^- (mg/l)	Ca^{2+} (mg/l)	Mg^{2+} (mg/l)	TA (mg/l)
S11	6.5	795	534.3	6.1	9.6	263.23	1.21	41	87.5	44.74	147.8
S12	6.8	633	493	8.1	7.8	260	0.3	53.96	96.03	20.21	200
S13	6.82	732	582	7.5	6.2	300	0.38	39.76	94.85	63.16	200
S14	6.8	793	590	7.6	6	360	0.35	53.96	120.64	58.76	190
S15	7.22	730	586	6.1	8	280	0.07	56.8	82.02	75.2	200
S16	7.06	635	405	6.9	10.2	320	0.33	48.28	92.04	90.18	150
S17	7.15	534	377	7.1	9.44	220	0.27	36.92	65.44	56.6	150
S18	7.06	453	232.8	7.1	11	280	0.01	52.54	81.22	77.19	100
S19	7.7	443	321.9	5.2	12	80	0.05	38.83	28.82	8.04	200
WHO drinking water quality standards [1]	6.5-8.5	500-1500	500-1000	-	-	-	50	250	75-200	30-150	-
CPCB Standards [2]	6.5-8.5	-	-	5	3	-	-	-	-	-	-

TABLE S-2
Targeted PAHs compounds, number of rings, molecular Weight & formula, retention time, LOD &LOQ in the instrument

PAH Compound	No. of Rings	Molecular		Retention Time	LOD (ng/l)	LOQ (ng/l)
		Weight	**Formula**			
Naphthalene (Nap)	2	128.18	$C_{10}H_8$	7.915	0.5	1.0
Acenaphthylene (Acpy)	3	152.2	$C_{12}H_8$	8.161	0.55	1.0
Acenaphthene (Ace)	3	154.2	$C_{12}H_{10}$	9.429	0.45	1.0
Fluorene (Fl)	3	166.23	$C_{13}H_{10}$	9.857	0.45	1.0
Phenanthrene (Phe)	3	178.24	$C_{14}H_{10}$	10.592	0.25	0.7
Anthracene (Ant)	3	178.24	$C_{14}H_{10}$	11.256	0.25	0.7
Fluoranthene (Fla)	4	202.25	$C_{16}H_{10}$	11.527	0.25	0.7
Pyrene (Pyr)	4	202.25	$C_{16}H_{10}$	12.491	0.25	0.7
Benzo (a) anthracene (BaA)	4	228.29	$C_{18}H_{12}$	12.644	0.25	0.5
Chrysene (Chr)	4	228.29	$C_{18}H_{12}$	13.246	0.20	0.5
Benzo (b) fluoranthene (BbF)	5	252.32	$C_{20}H_{12}$	16.543	0.20	0.5
Benzo (k) fluoranthene (BkF)	5	252.32	$C_{20}H_{12}$	18.141	0.20	0.5
Benzo (a) pyrene (BaP)	5	252.32	$C_{20}H_{12}$	20.606	0.20	0.5
Dibenzo (a,h) anthracene (DahA)	5	278.35	$C_{22}H_{14}$	22.844	0.20	0.5
Benzo (g,h,i) perylene (BghiP)	6	276.34	$C_{22}H_{12}$	23.97	0.20	0.5
Indeno (1,2,3-cd) pyrene (Ind)	6	276.34	$C_{22}H_{12}$	28.71	0.20	0.5

The limit of detection (LOD) was defined as the lowest concentration giving a signal-to-noise ratio (S/N) of at least 3 for the quantifier MRM transition. The limit of quantification (LOQ) was defined as the lowest calibration level meeting S/N ≥ 10, recovery of 70-120% and relative standard deviation (RSD) < 10%. Linearity and quantification of spiked samples were assessed using matrix-matched calibration standards at five levels (0.5, 1.0, 2.0, 5.0 and 10 ng/l). All calibration standards were within 70-120% accuracy. Calibration was linear for all PAHs, with coefficients of determination (r^2) ranging from 0.9777 to 0.9899. LODs ranged from 0.20 to 0.55 ng/l and LOQs from 0.5 to 1.0 ng/l.

Matrix-matched calibration standards were prepared by spiking the matrix blank extract with the solvent standards. The matrix blank-to-solvent standard volume ratio was held constant at every level, so the matrix effect was constant across the calibration range (0.5-10 ng/l). For the recovery study, 300 ml water samples were spiked with the standard mixture stock solution at 0.5 and 1.0 ng/l, then extracted, analysed, and quantified against the matrix-matched calibration curve. Because the extract was concentrated 10-fold (300 ml to a final volume of 30 ml), the concentrations injected were 5 and 10 ng/l, respectively.

No internal standard was added. Quantification used external, matrix-matched calibration (0.5-10 ng/l) prepared in blank matrix extract, which compensates for matrix effects on ionization. The 16 PAHs were quantified using a certified reference standard mixture (16 USEPA priority PAHs were quantified using an external certified reference standard mixture (Brand: Supalco, catalogue no. 49156 EPA TCL PAH Mix certified reference material, in acetonitrile: methanol (9:1), lot No LRAD7500). No internal standards were used.

TABLE S-3

MRM transitions for PAH quantification (GC-MS/MS)

Compound Name	RT (min)	Quantifier transition (m/z)	Qualifier transition (m/z)	Typical Collision Energy (CE)
Naphthalene	8.015	128.0	102.0 / 127.0	20
Acenaphthylene	8.361	152.0	150.0 / 126.0	28
Acenaphthene	9.129	154.0	152.0 / 153.0	28
Fluorene	9.957	166.0	165.0 / 152.0	28
Phenanthrene	10.892	178.0	152.0 / 176.0	25
Anthracene	11.056	178.0	152.0 / 176.0	20
Fluoranthene	11.327	202.0	101.0 / 200.0	28
Pyrene	12.091	202.0	101.0 / 200.0	30
Benz[a]anthracene	12.244	228.0	226.0 / 202.0	28
Chrysene	13.046	228.0	226.0 / 202.0	28
Benzo[b]fluoranthene	16.143	252.0	250.0 / 226.0	28
Benzo[k]fluoranthene	18.141	252.0	250.0 / 226.0	28
Benzo[a]pyrene	20.106	252.0	250.0 / 225.0	28
Indeno[1,2,3-cd]pyrene	21.844	276.0	274.0 / 250.0	28
Dibenz[a,h]anthracene	21.97	278.0	276.0 / 252.0	30
Benzo[ghi]perylene	28.71	276.0	274.0 / 250.0	30

TABLE S-4
Comparison of PAH concentrations in the present study and other Indian surface water

Study area	Water type	PAHs analysed	PAHs found	ΣPAHs concentration (ng/l)	Reference
Present study	Surface water	16 priority PAHs	Fluorene, phenanthrene, chrysene and benzo[a]pyrene	ND–10.61	Present study
Budhabalanga River Estuary, Chandipur, Odisha	Estuarine surface water	16 priority PAHs	Multiple PAHs; low-molecular-weight PAHs contributed substantially	19.5–1206	[3]
Mahi River Estuary, Gulf of Khambhat, Gujarat	Estuarine water	13 PAHs	Naphthalene, acenaphthylene, acenaphthene, fluorene, phenanthrene, anthracene, pyrene and other targeted PAHs	119–4648	[4]
Hindon River, Yamuna River Uttar Pradesh, India	River surface water	16 priority PAHs	Naphthalene, acenaphthylene, fluorene, and chrysene	Hindon River: 1640–73,400 Yamuna River: 1980–93,000	[5]
Hooghly River, eastern India and Brahmaputra River, northeastern India	River surface water	16 priority PAHs	Naphthalene, Acenaphthylene, Acenaphthene, Fluorene, Phenanthrene, Anthracene, Benzo(a)anthracene, Chrysene, Benzo (b)flouranthene, Benzo(k)flouranthene	Hooghly River, Mean: 4800 Brahmaputra River, Mean: 4040	[6]
Ganga River basin	River surface water	16 priority PAHs	Phenanthrene, Pyrene, Fluoranthene, Fluorine, Benzo(A)Anthracene	0.05–65.9	[7]

References

World Health Organization, Guidelines for Drinking-water Quality, fourth ed., incorporating the first addendum, WHO, Geneva, Switzerland, (2017).

Central Pollution Control Board (CPCB), Water Quality Criteria and Classification of Inland Surface Waters, Ministry of Environment, Forest and Climate Change, Government of India, New Delhi, (2019).

S. Nayak, S.P. Pradhan, P. Nayak, S.N. Sharma, N. Muduli, S. Nayak, L. Patnaik, Occurrence and risk assessment of polycyclic aromatic hydrocarbons in the surface water of Budhabalanga River Estuary, Chandipur, East Coast of India, Applied Biochemistry and Biotechnology 197 (2025) 4644–4664, https://doi.org/10.1007/s12010-025-05248-7.

T.P. Sahoo, P. Thathola, A.K. Panja, M. Bhowmik, B. Swain, S. Ray, H.T. Saravaia, Occurrence and distribution of organic pollutants for the assessment of ecological health status of Mahi River Estuary, Gulf of Khambhat, Journal of Hazardous Materials Advances 18 (2025) 100706, https://doi.org/10.1016/j.hazadv.2025.100706.

B. Kumar, V.K. Verma, S. Kumar, P. Gargava, Polycyclic aromatic hydrocarbons in surface waters from India: Possible sources and risk assessment, Polycyclic Aromatic Compounds 44 (2024) 3161–3177, https://doi.org/10.1080/10406638.2023.2231600

S.N. Khuman, P. Chakraborty, A. Cincinelli, D. Snow, B. Kumar, Polycyclic aromatic hydrocarbons in surface waters and riverine sediments of the Hooghly and Brahmaputra Rivers in the Eastern and Northeastern India, Science of the Total Environment 636 (2018) 751–760, https://doi.org/10.1016/j.scitotenv.2018.04.109.

B.M. Sharma, L. Melymuk, G.K. Bharat, P. Přibylová, O. Sáňka, J. Klánová, L. Nizzetto, Spatial gradients of polycyclic aromatic hydrocarbons (PAHs) in air, atmospheric deposition, and surface water of the Ganges River basin, Science of the Total Environment 627 (2018) 1495–1504, https://doi.org/10.1016/j.scitotenv.2018.01.262

TABLE S-5
Diagnostic Ratio of different sampling sites

Sampling Sites No.	Fl	Phe	Chr	BaP	ΣLMW	ΣHMW	ΣLMW/ΣHMW	Source indication
S1	10.6	ND	ND	ND	10.6	ND	NA	-
S2	2.7	ND	ND	ND	2.7	ND	NA	-
S3	1.6	ND	ND	ND	1.6	ND	NA	-
S4	1.8	ND	ND	ND	1.8	ND	NA	-
S5	2.5	ND	ND	ND	2.5	ND	NA	-
S6	ND	ND	ND	ND	ND	ND	NA	-
S7	2.7	ND	ND	ND	2.7	ND	NA	-
S8	6.1	ND	ND	ND	6.1	ND	NA	-
S9	3.5	ND	ND	ND	3.5	ND	NA	-
S10	ND	ND	ND	ND	ND	ND	NA	-
S11	ND	0.33	0.34	ND	0.33	0.34	**0.97**	**combustion**
S12	ND	0.88	ND	ND	0.88	ND	NA	-
S13	ND	2.95	1.44	ND	2.95	1.44	**2.05**	**Petrogenic**
S14	ND	ND	ND	3.25	ND	3.25	NA	-
S15	ND	ND	0.67	ND	ND	0.67	NA	-
S16	ND	ND	0.82	ND	ND	0.82	NA	-
S17	ND	ND	ND	ND	ND	ND	NA	-
S18	ND	ND	ND	ND	ND	ND	NA	-
S19	ND	ND	ND	ND	ND	ND	NA	-

TABLE S-6
The network information for proposed ANN (Fl and Chr)

Network Information of Fl			
Input Layer	**Covariates**	1	pH
		2	EC
		3	TDS
		4	DO
		5	BOD
		6	TH
		7	NO3
		8	Cl
		9	Ca
		10	Mg
		11	TA

		12	Phe
		13	Chr
		14	Bap
	Number of Units[a]		14
	Rescaling Method for Covariates		Standardized
	Number of Hidden Layers		1
Hidden Layer(s)	Number of Units in Hidden Layer 1[a]		8
	Activation Function		Hyperbolic tangent
	Dependent Variables	1	Fl
	Number of Units		1
Output Layer	Rescaling Method for Scale Dependents		Standardized
	Activation Function		Identity
	Error Function		Sum of Squares

a. Excluding the bias unit

Network Information of Chr			
		1	pH
		2	EC
		3	TDS
		4	DO
		5	BOD
		6	TH
	Covariates	7	NO3
		8	Cl
Input Layer		9	Ca
		10	Mg
		11	TA
		12	Fl
		13	Phe
		14	Bap
	Number of Units		14
	Rescaling Method for Covariates		Standardized
	Number of Hidden Layers		1
Hidden Layer(s)	Number of Units in Hidden Layer 1[a]		8
	Activation Function		Hyperbolic tangent
	Dependent Variables	1	**Chr**
	Number of Units		1
Output Layer	Rescaling Method for Scale Dependents		Standardized
	Activation Function		Identity
	Error Function		Sum of Squares

a. Excluding the bias unit

TABLE S-7
Sum of Squares Error and Relative error of Fl and Chr

		Sum of Squares Error	**Relative Error**
Fl	**Training**	0.005	0.001
	Testing	0.095	0.066
Chr	**Training**	0.529	0.081
	Testing	0.086	0.084

TABLE S-8
PMF modelling for source apportionment supported by field observations

Sampling site No.	Description	Field observations	PMF Modeling
S1	Pond near Tyre Dealer Centre Chhatimtala,	High Fl at the tyre dealer centre, Chhatimtala, Chakdaha: 10.6 ng/l	F3: Biomass/gaseous-fuel combustion and other low-Molecular Weight anthropogenic sources

Sampling site No.	Description	Field observations	PMF Modeling
	Chakdaha	indicates a low-molecular-weight PAH source.	
S2	Pond near Car washing centreChakdaha	Fl concentration at Pond near Car-washing centre, Chakdaha: 2.7 ng/l	F3: Biomass/gaseous-fuel combustion and other low-Molecular Weight anthropogenic sources
S3	Ghat of idol Immersion of Churni River, Ranaghat	Fl concentration at ghat of idol Immersion of Churni River 1.6 ng/l	F3: Biomass/gaseous-fuel combustion and other low-Molecular Weight anthropogenic sources
S4	Bathing Ghat of Churni River, Ranaghat	Fl and Chr concentration at Bathing Ghat of Churni River, Ranaghat: 1.8 ng/l and 0.18ng/l	F2: Petrogenic/anthropogenic sources associated with industrial and vehicular activities F3: Biomass/gaseous-fuel combustion and other low-Molecular Weight anthropogenic sources
S5	Pond near Highway NH-12 Chakdaha	Fl concentration at Pond near Highway NH-12 Chakdaha: 2.5 ng/l	F3: Biomass/gaseous-fuel combustion and other low-Molecular Weight anthropogenic sources
S6	Bikrampur ghat (Kulia jhill)	Fl concentration at Kuliya bill: 0.04 ng/l	F3: Biomass/gaseous-fuel combustion and other low-Molecular Weight anthropogenic sources
S7	Pond near TDK industry	Fl concentration at Pond near TDK industry: 2.7 ng/l	F3: Biomass/gaseous-fuel combustion and other low-Molecular Weight anthropogenic sources
S8	Kalyani lake	Fl concentration at Kalyani lake: 6.1 ng/l	F3: Biomass/gaseous-fuel combustion and other low-Molecular Weight anthropogenic sources
S9	Pond near Sati mayerbari	Fl concentration at pond near Sati mayerbari: 3.5 ng/l	F3: Biomass/gaseous-fuel combustion and other low-Molecular Weight anthropogenic sources
S10	Sundarpur (canal near river)	Fl concentration at Sundarpur (canal near river): 0.4 ng/l	F3: Biomass/gaseous-fuel combustion and other low-Molecular Weight anthropogenic sources
S11	Sundarpur (pond)	Phe and Chr concentration at Sundarpur pond: 0.33 ng/l and 0.34 ng/l	F1: Fossil-fuel combustion F2: Petrogenic sources associated with industrial and vehicular activities
S12	Halisahar ferry ghat	Phe concentration at Halisahar ferry ghat: 0.88 ng/l	F1: Fossil-fuel combustion
S13	Naihati ferry ghat	Fl, Phe and Chr concentration at Naihati ferry ghat: 0.1ng/l, 2.95 ng/l and 1.44 ng/l.	F1: Fossil-fuel combustion F2: Petrogenic sources associated with industrial and vehicular activities F3: Biomass/gaseous-fuel combustion and other low-Molecular Weight anthropogenic sources
S14	Shyamnagar ghat	BaP concentration at Shyamnagar ghat: 3.25 ng/l.	F3: Biomass/gaseous-fuel combustion and other anthropogenic sources
S15	Shyamnagar ferry ghat	Chr concentration at Shyamnagar ferry ghat: 0.67 ng/l	F2: Petrogenic/anthropogenic sources associated with industrial and vehicular activities
S16	North Barrackpore ferryghat	Phe and Chr concentration at North Barrackpore ferryghat: 0.11 ng/l and 0.82 ng/l	F1: Fossil-fuel combustion F2: Petrogenic sources associated with industrial and vehicular activities
S17	Kulia jhill	No priority PAHs have been found	-
S18	Dogachhina (road side lake)	No priority PAHs have been found	-
S19	Kalyani Block B (car washing)	No priority PAHs have been found	-

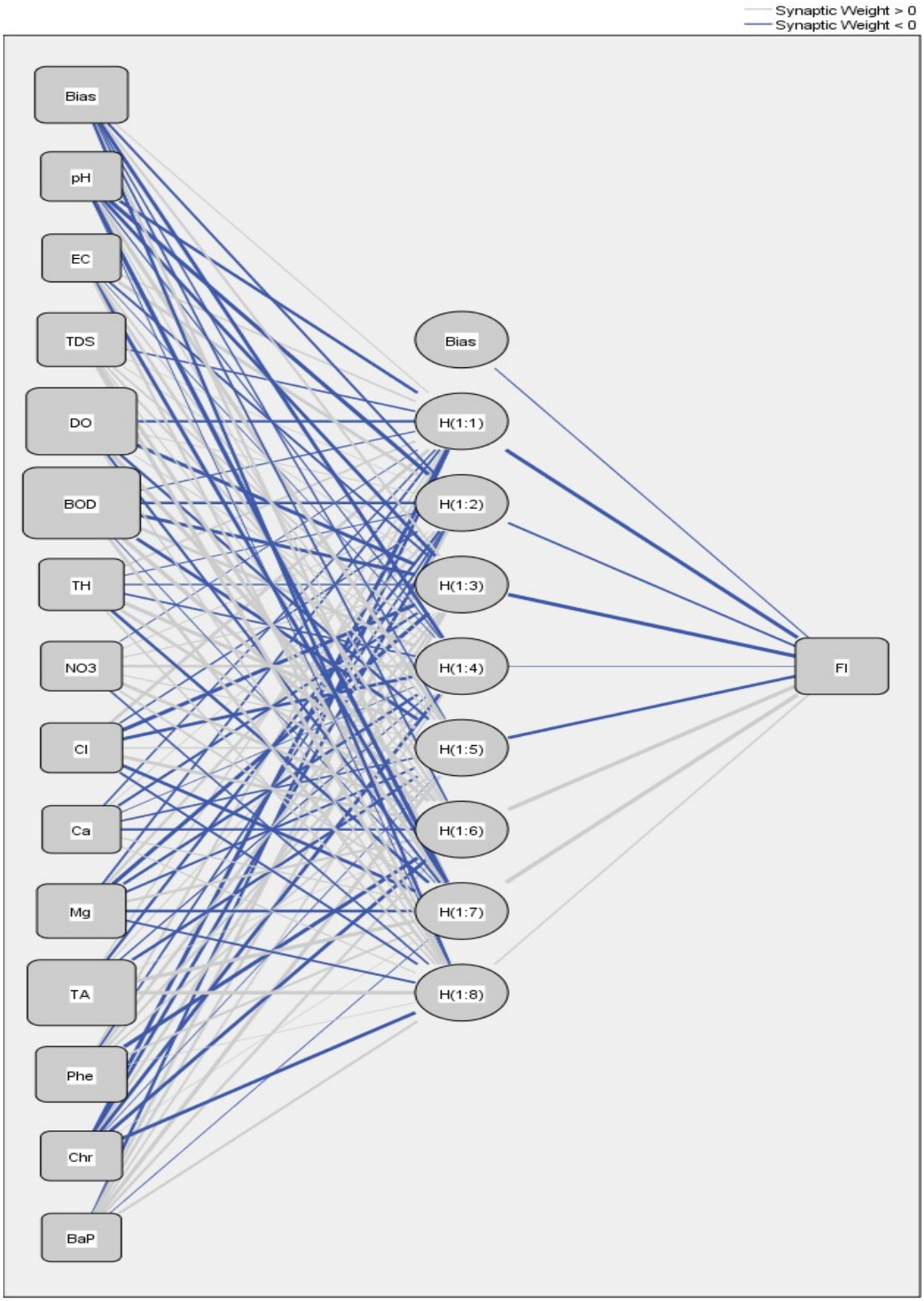


Fig. S-1: The implemented ANN of Fluorene (Fl)

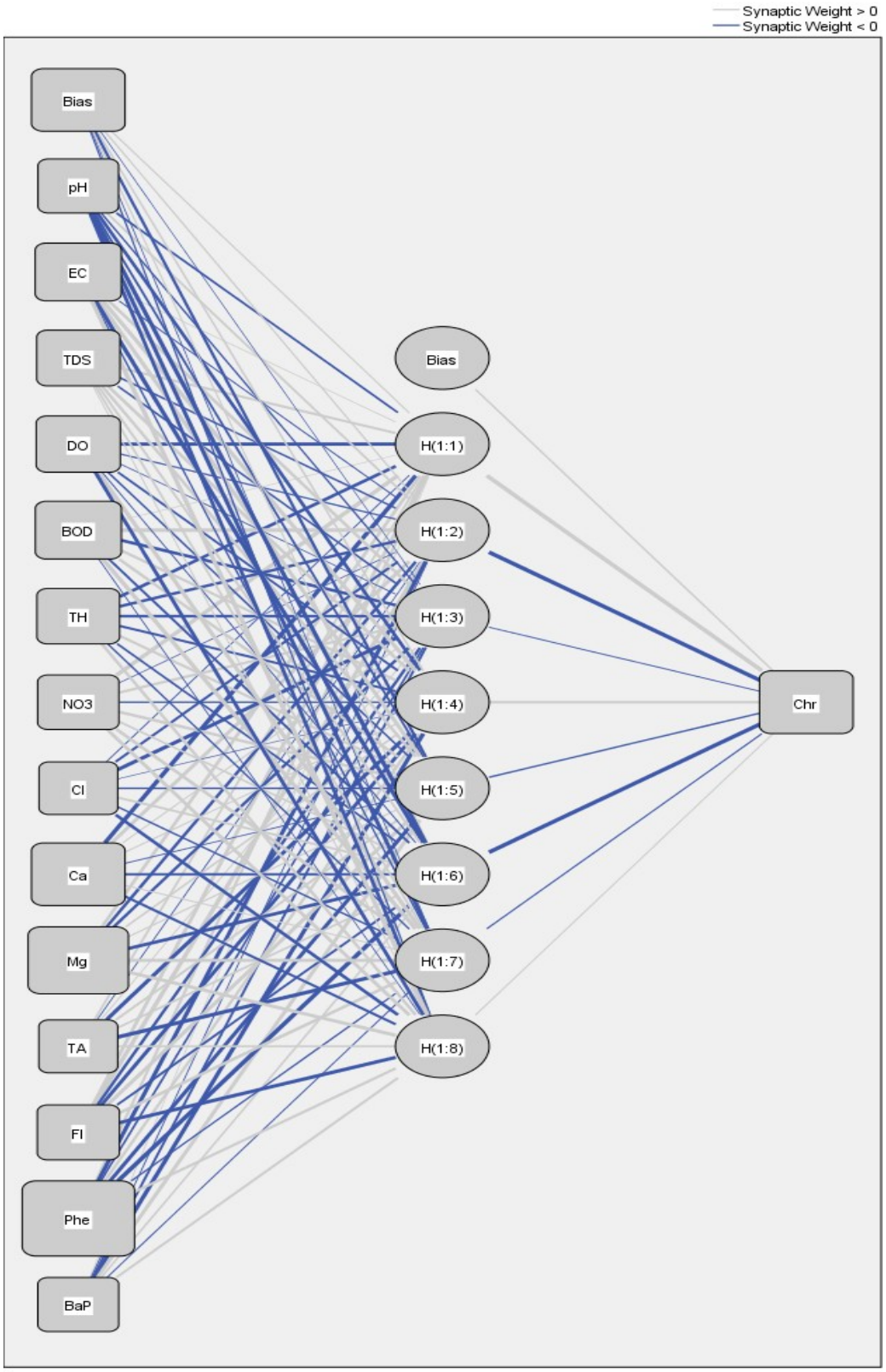


Fig. S-2: The implemented ANN of Chrysene (Chr)

X----------X ---------- X ----------X ----------X